

**10/563832**

**IAP20 RESU.PCTPID 09 JAN 2006**

**COURTESY COPY OF THE  
PCT APPLICATION AS  
ORIGINALLY FILED  
WITH SEQUENCE LISTING  
AND ABSTRACT**

## DIAGNOSIS AND TREATMENT METHODS RELATED TO AGING (8A)

This application claims the benefit, under 35 USC 119(e), of U.S. Provisional Appl. No. 60/485,222, filed July 8, 2003, which is hereby incorporated by reference in its entirety.

### Mention of Government Grant

Some of the work disclosed herein was funded by NIH grant #AG19899. Consequently, the federal government may enjoy certain rights in the invention.

### Cross-Reference to Related Applications

Anti-Aging Applications. Mice with a disrupted growth hormone receptor/binding protein gene enjoy an increased lifespan. In U.S. Prov. Appl. 60/485,222, filed July 8, 2003 (Kopchick8) mouse genes differentially expressed in comparisons of gene expression in growth hormone receptor/binding protein gene-disrupted mouse **livers** and normal mouse livers were identified, as were corresponding human genes and proteins. It was suggested that the human molecules, or antagonists thereof, could be used for protection against faster-than-normal biological aging, or to achieve slower-than-normal biological aging. It was also taught that the human molecules may also be used as markers of biological aging.

In provisional application Ser. No. 60/474,606, filed June 2, 2003 (our docket Kopchick7-USA) , our research group used a gene chip to study the genetic changes in the **liver** of C57Bl/6J mice that occur at frequent intervals of the aging process. Differential hybridization techniques were used to identify mouse genes that are differentially expressed in mice, depending upon their age. The level of gene expression of approximately 10,000 mouse genes (from the Amersham Codelink UniSet Mouse I Bioarray, product code: 300013) in the liver of mice with average ages of 35, 49, 56, 77, 118, 133, 207, 403, 558 and 725 days was determined. In essence, complementary RNA derived from mice of different ages was screened for hybridization with

oligonucleotide probes each specific to a particular mouse gene, each gene in turn representative of a particular mouse gene cluster (Unigene). Mouse genes which were differentially expressed (younger vs. older), as measured by different levels of hybridization of the respective cRNA samples with the particular probe corresponding to that mouse gene, were identified. Related human genes and proteins were identified by sequence comparisons to the mouse gene or protein. In the international appl.

Kopchick7A-PCT, filed June 2, 2004, we added some additional studies of CIDE-A (see below).

In a like manner, the effect of aging on the expression of genes in mouse skeletal muscle was studied, see provisional application Ser. No. 60/566,068, filed April 29, 2004 (our docket Kopchick14-USA).

*Anti-Diabetes Applications.* In U.S. Provisional Appl. Ser. No. 60/458,398 (our docket Kelder1-USA), filed March 31, 2003, members of our research group describe the identification of genes differentially expressed in normal vs. hyperinsulinemic, hyperinsulinemic vs. type II diabetic, or normal vs. type II diabetic mouse liver. Forward- and reverse-subtracted cDNA libraries were prepared, clones were isolated, and differentially expressed cDNA inserts were sequenced and compared with sequences in publicly available sequence databases. The corresponding mouse and human genes and proteins were identified.

The purpose of our research group's provisional application Ser. No. 60/460,415 (our docket: Kopchick6-USA), filed April 7, 2003, was similar, but complementary RNA, derived from RNA of mouse liver, was screened against a mouse gene chip. See also 60/506,716, filed Sept. 30, 2003 (Kopchick6.1).

Gene chip analyses have also been used to identify genes differentially expressed in normal vs. hyperinsulinemic, hyperinsulinemic vs. type II diabetic, or normal vs. type II diabetic mouse pancreas, see U.S. Provisional Appl. 60/517,376, filed Nov. 6, 2003 (Kopchick12) and muscle, see U.S. Provisional Appl.

3

60/547,512, filed Feb. 26, 2004 (Kopchick15).

*Other differential hybridization applications.* The use of differential hybridization to identify genes and proteins is also described in our research group's Ser. No.

PCT/US00/12145 (Kopchick 3A-PCT), Ser. No. PCT/US00/12366 (Kopchick4A-PCT), and Ser. No. 60/400,052 (Kopchick5).

All of the foregoing applications are hereby incorporated by reference in their entirety.

## BACKGROUND OF THE INVENTION

### Field of the Invention

The invention relates to various nucleic acid molecules and proteins, and their use in (1) diagnosing aging, or adverse conditions associated with the aging process, and (2) protecting mammals (including humans) against the aging process or adverse conditions associated with the aging process.

### Description of the Background Art

The mechanisms that cause aging (the decline in survival and reproductive ability with advancing age) have puzzled our society and scientific community for centuries. The two major theories center on the question of whether normal aging is an evolutionarily-genetically preprogrammed pathway of internal changes or is a normal consequence of existence where there is an accumulation of molecular and cellular damages. Hypotheses of such accumulated damage include free radical-oxidative damage, defective mitochondria, somatic mutations, progressive shortening of telomeres, programmed cell death, impaired cell proliferation and numerous others (1). The current belief is that aging is not a programmed process in that, to date, no genes are known to have evolved specifically to cause damage and aging. The one factor that has been shown to extend the lifespan in organisms from yeast to mice has been a reduction in caloric intake (2, 3). Recent data suggests that caloric restriction may also be relevant for primates,



including humans (4-6). Unfortunately, it is unlikely that most people will be able to maintain the strict dietary control required to reap the benefits of this finding.

Therefore, since the mechanism(s) by which caloric restriction extends lifespan are unknown, the elucidation of such mechanisms could lead to the development of alternative strategies to yield similar benefits.

Numerous groups are presently engaged in identifying genes and pathways that are involved in the aging process. A growing list of genes that extend adult longevity have been identified and a large proportion of these genes are involved with hormonal signals. Many of these genes and the corresponding endocrine systems are conserved among a wide variety of eukaryotes. What is becoming clear, at least in lower animal species, is that those pathways that provide advantages to development and growth early in life may impart negative consequences in later life. The clearest example of a genetic pathway affecting adult lifespan has been described in the nematode, *Caenorhabditis elegans*.

When food is abundant, *C. elegans* develops directly to the reproductive adult through four larval stages in three days. Under adverse conditions such as caloric restriction or high population density, *C. elegans* enters the Dauer diapause, a non-feeding, stress-resistant larval state. Genetic analysis has identified that mutation of single genes involved in dauer formation (*Daf*) greatly extend the adult lifespan (7). These genes involve the highly-conserved insulin/IGF-like signal transduction pathway. Ligand binding to the *daf-2* insulin-like receptor results in a kinase signaling cascade to phosphorylate the forkhead transcription factor, *daf-16*. This phosphorylation sequesters *daf-16* to the cytoplasm and results in reproductive maturity and aging. In the absence of ligand and signal transduction, the unphosphorylated, *daf-16* localizes to the nucleus and regulates the transcription of its target genes that promote dauer formation, stress resistance and extended longevity (8). A similar pathway has been described in *Drosophila melanogaster*. Mutation of the gene encoding insulin-like receptor (*InR*) or the gene

encoding insulin-receptor substrate (*chico*) also extends the normal life-span (9,10). Vertebrate homologues of *daf-16* down-regulate genes promoting cell progression, induce genes involved in DNA-damage repair and up-regulate genes that reduce intracellular reactive oxygen species (ROS) (11,12). A second *C. elegans* gene, *clk-1*, has also been linked to the reduction of ROS and an extended life-span. While the effect of *daf-2* mutants result in a reduction of mitochondrial ROS, *clk-1* mutants reduce extramitochondrially produced ROS. Since the majority of cellular ROS is produced in the mitochondria during the process of electron transport, it is not surprising that *clk-1* mutants have only a moderately extended life-span. *C. elegans* containing *daf-2/clk-1* double mutations, however, exhibit a very long life-span (13).

Decreased IGF-1 signaling may also extend longevity in mice. Four mouse models with deficiencies in pituitary endocrine action have demonstrated retarded aging. In the *Prop1* and *Pit1* models, pituitary production of growth hormone (GH), prolactin (PRL) and thyroid stimulating hormone (TSH) are ablated. These mice have reduced growth rates, reduced adult body size and live 40 to 60% longer than normal mice (14,15). Unfortunately, it is not possible to determine which of the ablated hormones is responsible for the increased longevity of the models.

A more straight-forward model was developed that targeted the deletion of the growth hormone receptor (*GHR-KO*) (16). This mouse line was derived from a founder animal by homologous recombination resulting in deletion and gene substitution of most of the fourth exon and part of the fourth intron of the *GHR/BP* gene. These mice also exhibit reduced body size and extended life-span and more directly implicates the GH/IGF-1 axis (17, 17a).

Recently, evidence for a direct role of IGF-1 receptor signaling in affecting the aging process was provided by the targeted disruption of the IGF-1 receptor (*Igf1r*) (18). Heterozygous females, but not males, possess 50% fewer receptors for IGF-1, live 33% longer than wild-type females and also display greater resistance to oxidative stress.

Tyrosine phosphorylation of the intracellular signaling molecule, *Shc*, was also decreased in the *Igf1r*<sup>+/-</sup> females. Mice containing the targeted deletion of p66shc also have increased resistance to oxidative stress and a 30% increase in life span (19). While the IGF-1 axis appears to be involved in the aging process, the mechanism by which it does so remains unknown. However, these findings demonstrate that it is possible to identify specific genetic pathways that affect the aging process. The finding that caloric restriction of these mouse models can further extend their life-span suggests that multiple pathways exist that affect the aging process (20). Therefore, research to identify these pathways and the genes involved in the aging process is of great importance.

#### ***Identification of genes involved in aging***

Several groups have begun to utilize DNA microarrays to measure differences in gene expression caused by the aging process. However, these experiments are extremely limited in regards to the number of aging time points or experimental conditions. For example, gene expression profiling has been performed on skeletal muscle tissue of mice at 5 versus 30 months of age with or without caloric restriction (21). In this analysis, the expression of 113 genes was found to be changed by at least two-fold in 5-month old mice compared to 30-month old mice. Caloric restriction of comparable mice caused a reversal of the altered gene expression of 33 genes. Similar analyses have also been performed on mouse brain and heart (22,23).

#### ***Differential Expression in Liver as a result of the aging process***

Cao, S.X., et al., "Genomic profiling of short- and long-term caloric restriction effects in the liver of aging mice", *Proc. Natl. Acad. Sci. USA*, 98:10630-10635 (2001) used Affymetrix microarray technology to study the changes in expression levels of 11,000 genes in liver tissue of 7 month-old mice compared to 27 month-old mice. In this

analysis, the expression of 20 genes increased at least 1.7-fold with age while the expression of 26 genes decreased at least 1.7-fold with age.

Tollet-Egnell, P., et al., "Gene expression profile of the aging process in rat liver: normalizing effects of growth hormone replacement, *Mol. Endocrinol.*, 15(2):308-18 (2001) used microarray technology to study the effect of aging and growth hormone treatment on the expression of 3,000 different genes in the rat liver. The proteins which were over-expressed in the older rat were glucose-6-phosphate isomerase (x1.8), pyruvate kinase (x4.8), hepatic product spot 14 (2.4x), fatty acid synthase (1.9x), staryl CoA desaturase (1.7x), enoyl CoA hydratase (1.7x), peroxisome proliferator activated receptor- $\alpha$  (1.7x), 3-ketoacyl-CoA thiolase (1.7x), 3-keto-acyl-CoA peroxisomal thiolase (1.9x), CYP4A3 (3.3x), glycerol-3-phosphate dehydrogenase (1.7x), NADPH-cytochrome P450 oxidoreductase (4.7x), CYP2C7 (1.9x), CYP3A2 (2.8x),  $\Delta$ -aminoevulinate synthase (2.3x). The under-expressed proteins were glucose-6-phosphatase (0.3x), farnesyl pyrophosphate synthase (0.5x), carnitine octanoyltransferase (0.5x), mitochondrial genome (16S ribosomal RNA) (0.3x), mitochondrial cytochrome c oxidase II (0.4x), mitochondrial NADH dehydrogenase SU 5 (0.3x), mitochondrial cytochrome b (0.4x), mitochondrial NADH dehydrogenase SU 3 (0.5x), NADH-ubiquinone oxidoreductase (SU CI-SGDH and SU 39kDa) (both 0.5x), ubiquinol-cytochrome c reductase (Rieske iron-sulfur protein and core 1) (both 0.5x), CYP2C12 (0.4x), cystathione  $\gamma$ -lyase (0.3x), biphenyl hydrolase-related protein (0.5x), glutathione S-transferase (class pi) (0.3x),  $\alpha$ -1 macroglobulin (0.5x), BRAK related protein (0.3x),  $\alpha$ -2u-globulin (0.4x), cAMP-dependent transcription factor mATF4 (0.5x), DAP-like kinase (0.5x), PCTAIRE-1 (0.5x), collagen  $\alpha$ -1' (0.4x), histone H2A (0.5x), and S-100 protein  $\alpha$  (0.5x).

#### *Differential/Subtractive Hybridization*

Zhang, et al., *Kidney International*, 56:549-558 (1999) identified genes up-regulated in 5/6 nephrectomized

(subtotal renal ablation) mouse kidney by a PCR-based subtraction method. Ten known and nine novel genes were identified. The ultimate goal was to identify genes involved in glomerular hyperfiltration and hypertrophy.

5 Melia, et al., *Endocrinol.*, 139:688-95 (1998) applied subtractive hybridization methods for the identification of androgen-regulated genes in mouse kidney. The treatment mice were dosed with dihydrotestosterone, an androgen. Kidney androgen-regulated protein gene was used as a  
10 positive control, as it is known to be up-regulated by DHT.

See also Holland, et al., Abstract 607, "Identification of Genes Possibly Involved in Nephropathy of Bovine Growth Hormone Transgenic Mice" (Endocrine Society Meeting, June 22, 2000) and Coschigano, et al., Abstract 333,  
15 "Identification of Genes Potentially Involved in Kidney Protection During Diabetes" (Endocrine Society Meeting, June 22, 2000).

The following differential hybridization articles may also be of interest: Wada, et al., "Gene expression  
20 profile in streptozotocin-induced diabetic mice kidneys undergoing glomerulosclerosis", *Kidney Int.*, 59:1363-73 (2001); Song, et al., "Cloning of a novel gene in the human kidney homologous to rat munc13S: its potential role in diabetic nephropathy", *Kidney Int.*, 53:1689-95 (1998); Page,  
25 et al., "Isolation of diabetes-associated kidney genes using differential display", *Biochem. Biophys. Res. Comm.*, 232:49-53 (1997); Peradi, "Subtractive hybridization claims: An efficient technique to detect overexpressed mRNAs in diabetic nephropathy," *Kidney Int.* 53:926-31 (1998);  
30 Condorelli, *EMBO J.*, 17:3858-66 (1998); See also WO00/66784 (differential hybridization screening for brown adipose tissue); PCT/US00/12366, filed May 5, 2000 (differential hybridization screening for liver).

35 For genes thought to have aging inhibitory activity, see generally International Longevity Center, Workshop Reports, "Longevity Genes: From Primitive Organisms to Humans," and "Is there an 'Anti-Aging' Medicine?".

Inohara, et al., EMBO J., 17: 2526-33 (1998) describes two proteins, CIDE-A and CIDE-B, which are believed to be cell death activators with homology to the 45kDa subunit of the DNA fragmentation factor (DFF). DFF is cleaved by caspase-3 during apoptosis. CIDE-A and -B activate apoptosis. The C-terminal region appears to contain the effector domain. Inohara et al. reported that CIDE-A was expressed in several human tissues, but not in the liver.

For the use of microarrays in the identification of aging-related genes, see Miller, J. Gerontol., 56A: B52-57 (2001); Lee et al., Science, 285 :1390-93 (1999) and Nature Genetics 25: 294-7 ( 2000) (bioarray study of changes in mouse cerebellum and neocortex to detect age-associated genes).

Patents of possible interest include the following:

Kojima, USP 5,000,188 (1991) (an apparatus for measuring the physiological age of a subject).

Lin, USP 6,303,768 (2001) ("Methuselah gene").

Lippman, USP 4,695,590 ("Method for retarding aging")

West, USP 6,368,789 (2002) ("Screening methods to identify inhibitors of telomerase activity")

Dimri, USP 5,795,728 (1998) ("Biomarkers of cell senescence")

Jia, USP 6,326,209 (2001) ("Measurement and quantification of 17 ketosteroid -sulfates as a biomarker of biological age")

Other articles of interest include Kayo, et al., Proc. nat. Acad. Sci. (USA) 98:5093-98 (2001); Han, et al., Mch. Ageing Dev. 115:157-74 (2000); Dozmorov, et al., J. gerontol. A Biol. Sci. Med. Sci. 56:B72-B80 (2001); Dozmorov, et al., Id., 57: B99-B108 (2002); Miller, et al., Mol. Endocrinol., 16: 2657-66 (2002).

## Apoptosis and CIDE-A

Apoptosis is a form of programmed cell death that occurs in an active and controlled manner to eliminate unwanted cells. Apoptotic cells undergo an orchestrated cascade of morphological changes such as membrane blebbing, nuclear shrinkage, chromatin condensation, and formation of apoptotic bodies which then undergo phagocytosis by neighboring cells. One of the hallmarks of cellular apoptosis is the cleavage of chromosomal DNA into discrete oligonucleosomal size fragments. This orderly removal of unwanted cells minimizes the release of cellular components that may affect neighboring tissue. In contrast, membrane rupture and release of cellular components during necrosis often leads to tissue inflammation.

The process of apoptosis is highly conserved and involves the activation of the caspase cascade. Cohen, GM. (1997) Caspases: the executioners of apoptosis. *Biochem. J.* 326:1-16; Budihardjo, I., Oliver, H., Lutter, M., Luo, X., Wang, X. (1999) Biochemical pathways of caspase activation during apoptosis. *Annu. Rev. Cell. Dev. Biol.* 15:269-290; Jacobson, M.D., Weil, M., Raff, M.C. (1997). Programmed cell death in animal development. *Cell* 88:347-354. Caspases are a family of serine proteases that are synthesized as inactive proenzymes. Their activation by apoptotic signals such as CD95 (Fas) death receptor activation or tumor necrosis factor results in the cleavage of specific target proteins and execution of the apoptotic program. Apoptosis may occur by either an extrinsic pathway involving the activation of cell surface death receptors (DR) or by an intrinsic mitochondrial pathway. Yoon, J-H. Gores G.J. (2002) Death receptor-mediated apoptosis and the liver. *J. Hepatology* 37:400-410.

These pathways are not mutually exclusive and some cell types require the activation of both pathways for maximal apoptotic signaling. In type-I cells, death receptor activation leads to the recruitment and activation of caspases-8/10 and the rapid cleavage and activation of caspase-3 in a mitochondrial-independent manner. Hepatocytes are members of the Type-II cells in which

mitochondria are essential for DR-mediated apoptosis

Scaffidi, C., Fulda, S., Srinivasan, A., Friesen, C., Li, F., Tomaselli, K.J., Debatin, K.M., Krammer, P.H., Peter, M.E. (1998) Two CD95 (APO-1/Fas) signaling pathways. *EMBO J.* 17:1675-1687. In this pathway, the pro-apoptotic protein Bid is truncated by activated caspases-8/10 and translocates to the mitochondria. Luo, X., Budihardjo, I., Zou, H., Slaughter, C., Wang, X. (1998) Bid, a Bcl2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors. *Cell* 94:481-490; Li, H., Zhu, H., Xu, C.J., Yuan, J. (1998) Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell* 94:491-501. This translocation leads to mitochondrial cytochrome c release and eventual activation of caspases-3 and 7 via cleavage by activated caspase-9.

One of the substrates for activated caspase-3 is the DNA fragmentation factor (DFF). DFF is composed of a 45 kDa regulatory subunit (DFF45) and a 40 kDa catalytic subunit (DFF40). Liu, X., Zou, H., Slaughter, C., Wang, X. (1997) DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. *Cell* 89:175-184. DFF45 cleavage by activated caspase-3 results in its dissociation from DFF40 and allows the caspase-activated DNase (CAD) activity of DFF40 to cleave chromosomal DNA into oligonucleosomal size fragments. Liu, X., Li, P., Widlak, P., Zou, H., Luo, X., Garrard, W.T., Wang, X. (1998) The 40-kDa subunit of DNA fragmentation factor induces DNA fragmentation and chromatin condensation during apoptosis. *Proc. Natl. Acad. Sci. USA.* 95:8461-8466; Halenbeck, R., MacDonald, H., Roulston, A., Chen, T.T., Conroy, L., Williams, L.T. (1998) CPAN, a human nuclease regulated by the caspase-sensitive inhibitor DFF45. *Curr Biol.* 8:537-540; Nagata, S. (2000) Apoptotic DNA fragmentation. *Exp. Cell Res.* 256:12-8.

Recently, a novel family of cell-death-inducing DFF45-like effectors (CIDes) have been identified that includes CIDE-A, CIDE-B and CIDE-3/FSP2. Inohara, N., Koseki, T., Chen, S., Wu, X., Nunez, G. (1998) CIDE, a novel family of



cell death activators with homology to the 45 kDa subunit of the DNA fragmentation factor. EMBO J. 17:2526-2533; Danesch, U., Hoeck, W., Ringold, G.M. (1992) Cloning and transcriptional regulation of a novel adipocyte-specific gene, FSP27. CAAT-enhancer-binding protein (C/EBP) and C/EBP-like proteins interact with sequences required for differentiation-dependent expression. J. Biol. Chem. 267:7185-7193; Liang, L., Zhao, M., Xu, Z., Yokoyama, K.K., Li, T. (2003) Molecular cloning and characterization of CIDE-3, a novel member of the cell-death-inducing DNA-fragmentation-factor (DFF45)-like effector family. Biochem. J. 370:195-203.

The CIDEs contain an N-terminal domain that shares homology with the N-terminal region of DFF45 and may represent a regulatory region via protein interaction. See Inohara, supra; Lugovskoy, A.A., Zhou, P., Chou, J.J., McCarty, J.S., Li, P., Wagner, G. (1999) Solution structure of the CIDE-N domain of CIDE-B and a model for CIDE-N/CIDE-N interactions in the DNA fragmentation pathway of apoptosis. Cell 9:747-755. The family members also share a C-terminal domain that is necessary and sufficient for inducing cell death and DNA fragmentation; see Inohara supra. The overexpression of CIDE-A induces cell death that can be inhibited by DFF45. However, CIDE-A-induced apoptosis is not inhibited by caspase-8 inhibitors thereby suggesting the presence of additional, caspase-independent, pathway(s) for the induction of apoptosis, see Inohara supra. Previous reports have indicated that human and mouse CIDE-A are expressed in several tissues such as brown adipose tissue (BAT) and heart and are localized to the mitochondria, Zhou, Z., Yon Toh, S., Chen, Z., Guo, K., Ng, C.P., Ponniah, S., Lin, S.C., Hong, W., Li, P. (2003) Cidea-deficient mice have lean phenotype and are resistant to obesity. Nat. Genet. 35:49-56. . In addition to the ability to induce apoptosis, CIDE-A can interact and inhibit UCP1 in BAT and may therefore play a role in regulating energy balance, see Zhou supra.

Previous reports have indicated that CIDE-A is not expressed in either adult human or mouse liver tissue, see Inohara supra, Zhou supra.

The human protein cell death activator CIDE-A is of particular interest because of its highly dramatic change in liver expression with age, first demonstrated in our  
5 Kopchick7 application, supra. CIDE-A expression is elevated in older normal mice. CIDE-A expression was studied for normal C57BI/6J mouse ages 35, 49, 77, 133, 207, 403 and 558 days. Expression is low at the first five data points, then rises sharply at 403 days, and again at 558 days.

10 CIDE-A was therefore classified as an "unfavorable protein", i.e., it was taught that an antagonist to CIDE-A could retard biological aging.

In Kopchick7A-PCT we reported that CIDE-A is also prematurely expressed in hyperinsulinemic and type-II  
15 diabetic mouse liver tissue. CIDE-A expression also correlates with liver steatosis in diet-induced obesity, hyperinsulinemia and type-II diabetes. These observations suggest an additional pathway of apoptotic cell death in Non-Alcoholic Fatty Liver Disease (NAFLD) and that CIDE-A  
20 may play a role in this serious disease and potentially in liver dysfunction associated with type-II diabetes.

## SUMMARY OF THE INVENTION

Our research group has previously reported that mice with a disrupted growth hormone receptor/binding protein gene enjoy an increased lifespan, as shown in the table below (17a):

Bkgd	Sex	Type	Mean Lifespan	Median Lifespan	% Surv >1000d
C57BL/6j	M	C	756 ±68	866	5
	M	KO	*951 ±50	941	43
	F	C	821 ±49	850	18
	F	KO	†956 ±80	1023	63
Ola-BALB/cJ	M	C	656 ±67	698	0
	M	KO	**917 ±55	888	27
	F	C	759 ±41	765	13
	F	KO	**921 ±41	981	43

\* sig at 0.05 level. \*\* sig at 0.01 level † not sig, but two KO females still alive at time of analysis and thus had not reached maximum lifespan, affecting the analysis. KO=knockout (GHR/BP -/-) mice, C=control (GHR/BP +/+) mice.

Our attention recently has focused on the generation of liver mRNA expression profiles and the identification of genes involved in the aging process. We have sought to identify genes and proteins which affect aging by determining which mouse genes are differentially expressed when expression in GHR/BP knockout mice and control mice is compared. Related human genes and proteins may then be identified e.g., by comparison of the mouse sequences with human sequences in sequence databanks.

After identifying related human genes and proteins, one may formulate agents useful in estimating the biological age of a human subject or in predicting the rate of biological aging in a human subject. It is particularly desirable to screen for individuals at risk for faster-than-normal biological aging.

Thus, "favorable" human genes/proteins are defined as those corresponding to mouse cDNAs which were less strongly expressed in control mouse liver than in knockout mice (longer lifespan) liver. Likewise, one may define "unfavorable" human genes/proteins as those corresponding to mouse cDNAs which were more strongly expressed in control mouse liver than in knockout mice (longer lifespan) liver.

As used herein, the term "corresponding" does not mean identical, but rather implies the existence of a statistically significant sequence similarity, such as one sufficient to qualify the human protein or gene as a homologous protein or DNA as defined below. The greater the degree of relationship as thus defined (i.e., by the statistical significance of each alignment used to connect the mouse cDNA to the human protein or gene, measured by an E value), the more close the correspondence. The connection may be direct (mouse cDNA to human protein) or indirect (e.g., mouse cDNA to mouse gene, mouse gene to human gene, human gene to human protein).

In general, the human genes/proteins which most closely correspond, directly or indirectly, to the mouse cDNA are preferred, such as the one(s) with the highest, top two highest, top three highest, top four highest, top five highest, and top ten highest E values for the final alignment in the connection process. The human genes/proteins deemed to correspond to our mouse cDNA clones are identified in the Master Tables.

Agents which bind the "favorable" and "unfavorable" nucleic acids (e.g., the agent is a substantially complementary nucleic acid hybridization probe), or the corresponding proteins (e.g., an antibody vs. the protein) may be used to evaluate whether a human subject is at increased or decreased risk for faster-than-normal biological aging. A subject with one or more elevated "unfavorable" and/or one or more depressed "favorable" genes/proteins is at increased risk, and one with one or more elevated "favorable" and/or one or more depressed "unfavorable" genes/proteins is at decreased risk.

The assay may be used as a preliminary screening assay to select subjects for further analysis, or as a formal diagnostic assay.

5           The identification of the related genes and proteins may also be useful in protecting humans against faster-than-normal or even normal aging (hereinafter, "the disorders").

          Thus, Applicants contemplate:

10           (1) use of the "favorable" mouse DNAs (cDNAs or genes) of the Master Tables (below) to isolate or identify related human DNAs;

          (2) use of human DNAs, related to favorable mouse DNAs, to express the corresponding human proteins;

15           (3) use of the corresponding human proteins (and mouse proteins, if biologically active in humans), to protect against the disorder(s);

          (4) use of the corresponding mouse or human proteins, or nucleic acid probes derived from the mouse or human  
20           genes, in diagnostic agents, in assays to measure biological aging or the rate thereof; and

          (5) use of the corresponding human or mouse genes therapeutically in gene therapy, to protect against the disorder(s).

25           Moreover Applicants contemplate:

          (1) use of the "unfavorable" mouse DNAs of the Master Tables to isolate or identify related human DNAs;

          (2) use of the complement to the "unfavorable" mouse  
30           DNAs or related human DNAs, as antisense molecules to inhibit expression of the related human DNAs;

          (3) use of the mouse or human DNAs to express the corresponding mouse or human proteins;

          (4) use of the corresponding mouse or human proteins,  
35           in diagnostic agents, to measure biological aging or the rate thereof;

          (5) use of the corresponding mouse or human proteins in assays to determine whether a substance binds to (and hence may neutralize) the protein; and

(6) use of the neutralizing substance to protect against the disorder(s).

5       The related human DNAs may be identified by comparing the mouse sequence (or its AA translation product) to known human DNAs (and their AA translation products). If this is unsuccessful, human cDNA or genomic DNA libraries may be screened using the mouse DNA as a probe.

10       If there is no closely corresponding full-length mouse gene in the sequence databank, and the cDNA is not full-length, then the mouse cDNA may be used as a hybridization probe to screen a mouse cDNA library to isolate the corresponding full-length sequence. Alternatively, the  
15       mouse cDNA may be used as a probe to screen a mouse genomic DNA library.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS OF THE INVENTION

### Subjects

5 For mice, infancy is defined as the period 0 to 21 days after birth. Sexual maturity is reached, on average, at 42 days after birth. The average lifespan is 832 days.

10 In humans, infancy is defined as the period between birth and two years of age. Sexual maturity in males can occur between 9 and 14 years of age while the average age at first menstrual period for females is 12.6 years. The average human lifespan is 73 years for males and 79 years for females. The maximum verified human lifespan was 122 years, five months and 14 days.

### Chronological and Biological Aging

15 "Aging" is a process of gradual and spontaneous change, resulting in maturation through childhood, puberty, and young adulthood and then primarily a decline in function through middle and late age. Aging thus has both the positive component of development/maturation and the negative component of decline.

20 "Senescence" refers strictly to the undesirable changes that occur as a result of post-maturation aging. Some of the changes which occur in post-maturation aging are not deleterious to health (e.g., gray hair, baldness), and some may even be desirable (e.g., increased wisdom and experience). In contrast, the memory impairment that occurs with age is considered senescence. However, we will hereafter use "aging" per se to refer to "senescence", and use "maturation" to refer to pre-maturation development.

25 There is increased mortality with age after maturation. There is also a progressive decrease in physiological capacity with age, but the rate of physiological decline varies from organ to organ and from individual to individual. The physiological decline results in a reduced ability to respond adaptively to environmental stimuli, and increased susceptibility and vulnerability to disease.

"Aging is the accumulation of diverse adverse changes that increase the risk of death. These changes can be attributed to development, genetic defects, the environment, disease, and the inborn aging process. The chance of death at a given age serves as a measure of the number of accumulated changes, that is, of physiologic age, and the rate of change of this measure, as the rate of aging." Harman, Ann. N.Y. Acad. Sci. 854:1-7 (1998).

Preferably, the agents of the present invention inhibit aging for at least a subpopulation of mature (post-puberty) adult subjects.

The term "healthy aging" (sometimes called "successful aging") refers to post-maturation changes in the body that occur with increasing age even in the absence of an overt disease. However, increased age is a risk factor for many diseases ("age-related diseases"), and hence "total aging" includes both the basal effects of healthy aging and the effects of any age-related disease. (Most literature uses the term "normal aging" as a synonym for "healthy aging", but a minority use it to refer to "total aging". To minimize confusion, we will try to avoid the term "normal aging", but if we use it, it is as a synonym for "healthy aging".) Some scientists have suggested that normal aging changes should be defined as those which are universal, degenerative, progressive and intrinsic.

Preferably, the agents of the present invention inhibit healthy aging for at least a subpopulation of mature (post-puberty) adult subjects.

In both aging and senescence, many physiologic functions decline, but normal decline is not usually considered the same as disease. The distinction between normal decline and disease is often but not always clear and may be due only to statistical distribution. Glucose intolerance is considered consistent with healthy aging, but diabetes is considered a disease, although a very common one. Cognitive decline is nearly universal with advanced age and is considered healthy aging; however, cognitive decline consistent with dementia, although common in late life, is considered a disease (as in the case of Alzheimer's, a



conclusion supported by analysis of brain tissue at autopsy). A decline in maximal heart rate is typical of healthy aging. In contrast, coronary heart disease is an age-related disease. A decline in bone density is considered healthy aging, but when it drops to 2.5 SD below the young adult mean, it is called osteoporosis. Generally speaking, the changes typical of healthy aging are gradual, while those typical of a disorder can be rapid.

The term average (median) "lifespan" is the chronological age to which 50% of a given population survive. The maximum lifespan potential is the maximum age achievable by a member of the population. As a practical matter, it is estimated as the age reached by the longest lived member (or former member) of the population. The (average) life expectancy is the number of remaining years that an individual of a given age can expect to live, based on the average remaining lifespans of a group of matched individuals.

The most widely accepted method of measuring the rate of aging is by reference to the average or the maximum lifespan. If a drug treatment achieves a statistically significant improvement in average or maximum lifespan in the treatment group over the control group, then it is inferred that the rate of aging was retarded in the treatment group. Similarly, one can compare long-term survival between the two groups.

Preferably, the agents of the present invention have the effect of increasing the average lifespan and/or the maximum lifespan for at least a subpopulation of mature (post-puberty) adult subjects. This subpopulation may be defined by sex and/or age. If defined in part by age, then it may be defined by a minimum age (e.g., at least 30, at least 40, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 90, etc.) or by a maximum age (not more than 40, not more than 50, not more than 55, not more than 60, not more than 65, not more than 70, not more than 75, not more than 80, not more than 90, not more than 100, etc.), or by a rational

combination of a minimum age and a maximum age so as to define a preferred close-ended age range, e.g., 55-75.

The subpopulation may additionally be defined by race, e.g., caucasian, negroid or oriental, and/or by ethnic group, and/or by place of residence (e.g., North America, Europe).

The subpopulation may additionally be defined by non-age risk factors for age-associated diseases, e.g., by blood pressure, body mass index, etc.

Preferably, the subpopulation in which an agent of the present invention is reasonably expected to be effective is large, e.g., in the United States, preferably at least 100,000 individuals, more preferably at least 1,000,000 individuals, still more preferably at least 10,000,000, even more preferably at least 20,000,000, most preferably at least 40,000,000.

By way of comparison, according to the 2000 U.S. Census, the U.S. population, by age, was

Age	Pop (mil)
15-19	20.2
20-24	19.0
25-29	19.4
30-34	20.5
35-39	22.7
40-44	22.4
45-49	20.1
50-54	17.6
55-59	13.5
60-64	10.8
65-69	9.5
70-74	8.9
75-79	7.4
80-84	4.9
85+	4.2

For any given chronological age, statisticians can define the probability of living to a particular later age. These expectancies can be calculated for the entire age cohort, or broken down by sex, race, country of residence, etc. Individuals who live longer than expected can be said, after the fact, to have biologically aged more slowly than their peers. One definition of biological age is that it is a measure of one's position in one's life span, i.e., biological age = position in own life span (as fraction in range 0..1) X average life span for species. This simple definition carries with it the implicit assumption that the rate of biological aging is constant. It also has the practical problem of determining one's own life span before death. We will present a more practical definition shortly.

The problem with lifespan studies is that they are extremely time-consuming. A maximum lifespan study in mice can take 4-5 years. A maximum lifespan study in dogs or cats would take 15-20 years, in monkeys, 30-40 years, and in humans, over 100 years. Even if the human study group were

of sexagenarians, it would take 40-60 years to complete the study.

Hence, scientists have sought to identify biological markers (biomarkers) of biological aging, that is, characteristics that can be measured while the subjects are still alive, which correlate to lifespan. These biological markers can be used to calculate a "biological age" (syn. "Physiological age"); it is the chronological age at which an average member of the population (or relevant subpopulation) would have the same value of a biomarker of biological aging (or the same value of a composite measure of biomarkers of biological aging) as does the subject. This is the definition that will be used in this disclosure, unless otherwise stated.

The effect of aging varies from system to system, organ to organ, etc. For example, between ages 30 and 70 years, nerve conduction velocity decreases by only about 10%, but renal function decreases on average by nearly 40%. Thus, there isn't just one biological age for a subject. By a suitable choice of biomarker, one may obtain a whole organism, or a system-, organ- or tissue-specific measure of biological aging, e.g., one can say that a person has the nervous system of a 30 year old but the renal system of a 60 year old. Biomarkers may measure changes at the molecular, cellular, tissue, organ, system or whole organism levels.

Generally speaking, in the absence of some form of intervention (drugs, diet, exercise, etc.), biological ages will increase with time. The agents of the present invention preferably reduce the time rate of change of a biological age of the subject. The term "a biological age" could refer to the overall biological age of the subject, to the biological age of a particular system, organ or tissue of that subject, or to some combination of the foregoing. More preferably, the agents of the present invention cannot only reduce the rate of increase of a biological age of the subject, but can actually reduce a biological age of the subject.

A simple biologic marker (biomarker) is a single biochemical, cellular, structural or functional indicator of an event in a biologic system or sample. A composite biomarker is a mathematical combination of two or more simple biomarkers. (Chronological age may be one of the components of a composite biomarker.)

A plausible biomarker of biological age would be a biomarker which shows a cross-sectional and/or longitudinal correlation with chronological age. Nakamura suggests that it is desirable that a biomarker show (a) significant cross-sectional correlation with chronological age, (b) significant longitudinal change in the same direction as the cross-sectional correlation, (c) significant stability of individual differences, and (d) rate of age-related change proportional to differences in life span among related species. Cp. Nakamura, *Exp Gerontol.* 29(2):151-77 (1994), using desiderata (a)-(c). A superior biomarker of biological age would be a better predictor of lifespan than is chronological age (preferably for a chronological age at which 90% of the population is still alive).

The biomarker preferably also satisfies one or more of the following desiderata: a statistically significant age-related change is apparent in humans after a period of at most a few years; not affected dramatically by physical conditioning (e.g., exercise), diet, and drug therapy (unless it is possible to discount these confounding influences, e.g., by reference to a second marker which measures them); can be tested repeatedly without harming the subject; works in lab animals as well as humans; simple and inexpensive to use; does not alter the result of subsequent tests for other biomarkers if it is to be used in conjunction with them; monitors a basic process that underlies the aging process, not the effects of disease.

Preferably, if the biomarker works in lab animals, there is a statistically significant difference in the value of the biomarker between groups of food-restricted and normally-fed animals. It has been shown in some mammalian species that dietary restriction without malnutrition (e.g.,

caloric decrease of up to 40% from ad libitum feeding) increases lifespan.

A biomarker of aging may be used to predict, instead of lifespan, the "Healthy Active Life Expectancy" (HALE) or the "Quality Adjusted Life Years" (QALY), or a similar measure which takes into account the quality of life before death as well as the time of death itself. For HALE, see Jagger, in *Outcomes Assessment for Healthcare in Elderly People*, 67-76 (Farrand Press: 1997). For QALY, see Rosser RM. A health index and output measure, in Stewart SR and Rosser RM (eds) *Quality of Life: Assessment and Application*. Lancaster: MTP, 1988.

A biomarker of aging may be used to predict, instead of lifespan, the timing and/or severity of a change in one or more age-related phenotypes as described below.

A biomarker of aging may be used to estimate, rather than overall biological age for a subject, a biological age for a specific body system or organ. The determination of the biological age of the liver, and the inhibition of biological aging of the liver, are of particular interest.

Body systems include the nervous system (including the brain, the sensory organs, and the sense receptors of the skin), the cardiovascular system (includes the heart, the red blood cells and the reticuloendothelial system), the respiratory system, the gastrointestinal system, the endocrine system (pituitary, thyroid, parathyroid and adrenal glands, gonads, pancreas, and paraganglia), the musculoskeletal system, the urinary system (kidneys, bladder, ureters, urethra), the reproductive system and the immune system (bone marrow, thymus, lymph nodes, spleen, lymphoid tissue, white blood cells, and immunoglobulins). A biomarker may be useful in estimating the biological age of a system because the biomarker is a chemical produced by that system, because it is a chemical whose activity is primarily exerted within that system, because it is indicative of the morphological character or functional activity of that system, etc. A given biomarker may be thus associated with more than one system. In a like manner, a

biomarker may be associated with the biological age, and hence the state, of a particular organ or tissue.

The prediction of lifespan, or of duration of system or organ function at or above a particular desired level, may require knowledge of the value of at least one biomarker of aging at two or more times, adequately spaced, rather than of the value at a single time. See McClearn, Biomarkers of Age and Aging, Exp. Gerontol., 32:87-94 (1997).

The levels (or changes in levels) of the human proteins identified in this specification, and their corresponding mRNAs, may be used as simple biomarkers (direct or inverse) of biological aging. They may be used in conjunction with each other, or other simple biomarkers, in a composite biomarker.

Once several plausible simple biomarkers have been identified, a composite biomarker may be obtained by standard mathematical techniques, such as multiple regression, principal component analysis, cluster analysis, neural net analysis, and so forth. As a preliminary to such analysis, the values may be standardized, e.g., by converting the raw scores into z-scores based on the distributions for each simple biomarker.

For example, principal component analysis can be used to analyze the variation of lifespan with different observables, and the factor score coefficients from the first principal component can be used to derive an equation for estimating a biological age score. Nakamura, Exp Gerontol. 29(2):151-77 (1994). This approach was used to obtain the following BAS (for healthy Japanese women aged 28-80):  $BAS = -4.37 - 0.998FEV_{1.0} + 0.022SBP + 0.133MCH + 0.018GLU - 1.505 A/G \text{ RATIO}$ , where  $FEV_{1.0}$  is the forced expiratory volume in 1 sec. (Liters), SBP is the systolic blood pressure (mm Hg), MCH is the mean corpuscular hemoglobin (pg), GLU is glucose (mg/dl), and A/G RATIO is the ratio of albumin to globulin. The relative importance of these five biomarkers was 33.7%, 25.1%, 17.1%, 14.8% and 8.9%, respectively. Ueno, et al., "Biomarkers of Aging in Women

and the Rate of Longitudinal Changes," J. Physiol. Anthropol. 22(1): 37-46 (Jan. 2003).

It should be noted that particularly when evaluating the overall biological age of the subject, it is not necessarily most desirable to weight all systems or all organs equally. One may find it more desirable to give greater weight to the system or organ with the highest biological age in calculating the overall biological age, because it is presumably more likely to deteriorate or fail, resulting in death. Appropriate statistical analysis can be used to find the weighting scheme resulting in the best prediction of lifespan.

In the H-SCAN (Hoch Company) test, a composite of 12 simple biomarkers is used to measure human aging:

#### SENSORY

1. Highest audible pitch (kHz)
2. Visual accommodation (diopters)
3. Vibrotactile sensitivity (dB)

#### MOTOR

4. Muscle Movement time (sec)
5. Muscle Movement time with decision (sec)
6. Alternate button tapping time (sec)

#### COGNITIVE

7. Memory, length of sequence
8. Auditory reaction time (sec)
9. Visual reaction time (sec)
10. Visual Reaction time with decision (sec)

#### PULMONARY

11. Forced vital capacity (liters)
12. Forced expiratory Volume- 1 sec (liters)

See Hochschild, R., Journal of Gerontology [Biological Science] 45(6):B187-214; 1990).



According to a website discussing the H-SCAN test,  
"Biomarkers of aging are characteristics of an organism that  
correlate in large groups with chronological age and  
mortality. Of particular value in human applications are  
5 biomarkers of aging that also correlate with the quality of  
life in later life in the sense that they involve functions  
that are crucial to carrying out the activities of daily  
living.... A single biomarker of aging is limited by the  
fact that it measures only one isolated characteristic and  
10 is hardly representative of the diversity of functional and  
structural concomitants of aging.... Biological age, in  
contrast to chronological age, is an individual's  
hypothetical age calculated from scores obtained on a  
battery of tests of biomarkers of aging. As a first step in  
15 the calculation, the age of which each biomarker score is  
typical is determined by comparison with scores obtained by  
a large representative group of persons (or organisms)  
spanning a range of ages. Then one of a variety of averaging  
techniques is employed (optionally with standardization  
20 steps) to obtain a single index of age, as described in  
detail by Hochschild. This index varies with, and therefore  
must be expressed with reference to, the measured biomarkers  
and the mathematical method of combining scores."

<http://www.longevityinstituteone.com/>

25 Abbo, USP 6,547,729 teaches determining the biological  
age (he calls it "performance age") of a subject by (1) for  
a sample population, determining a regression curve relating  
some set of observed values for an "indicator" of the  
30 functionality of a bodily system to the chronological age of  
the observed individuals, (2) solving the regression  
equation to obtain a predicted performance age, given the  
value of the indicator for the subject. The regression can  
be based on more than one indicator, i.e., it can be a  
35 multiple regression. The sample population can be defined  
by sex, age range, ethnic composition, and geographic  
location. The bodily system may be a molecular, cellular,  
tissue or organ system. The following indicators are  
suggested by Abbo: nervous system (memory tests, reaction

time, serial key tapping, digit recall test, letter fluency, category fluency, nerve conduction velocity), arteries (pulse wave velocity; ankle-brachial index), skeletal system (bone mineral density); lungs (forced vital capacity), heart (ejection fraction; length of time completed on a treadmill stress test), kidneys (creatinine clearance), proteins (glycosylation of hemoglobin), endocrine glands (load level of bioactive testosterone; level of dehydroepiandrosterone sulfate, ratio of urinary 17-ketosteroids/17-hydroxycorticosteroids; growth hormone; IGF-1).

Preferably, the agents of the invention have a favorable effect on the value of at least one simple biomarker of biological aging, such as any of the plausible biomarkers mentioned anywhere in this specification, other than the level of one of the proteins of the present invention. More preferably, they have a favorable effect on the value of at least two such simple biomarkers of biological aging. Even more preferably, at least one such pair is of markers which are substantially non-correlated ( $R^2 < 0.5$ ).

Desirably, if more than one simple biomarker is favorably affected, the biomarkers in question reflect different levels of organization, and/or different body components at the same level of organization. For example, a visual reaction time with decision test is on the whole organism level, while a measurement of telomere length is on the cellular level.

A biomarker may, but need not, be an indicator related to one of the postulated causes or contributing factors of aging. It may, but need not, be an indicator of the acute health of a particular body system or organ.

A biomarker may measure behavior, cognitive or sensory function, or motor activity, or some combination thereof. It may measure the level of a type of cell (e.g., a T cell subset, such as CD4, CD4 memory, CD4 naive, and CD4 cells expressing P-glycoprotein) or of a particular molecule

(e.g., growth hormone, IGF-1, insulin, DHEAS, an elongation factor, melatonin) or family of structurally or functionally related molecules in a particular body fluid (especially blood) or tissue. For example, lower serum IGF-1 levels are correlated with increasing age, and IGF-1 is produced by many different tissues. On the other hand, growth hormone is produced by the pituitary gland.

A biomarker may measure an indicator of stress (particularly oxidative stress) and resistance thereto. It has been theorized that free radicals damage biomolecules, leading to aging.

A biomarker may measure protein glycation or other protein modification (e.g., collagen crosslinking). It has been theorized that such modifications contribute to aging.

The biomarker may measure changes in the lengths of telomeres or in the rate of cell division. It has been theorized that telomere shortening beyond a critical length leads the cell to stop proliferating. Average telomere length therefore provides a biomarker as to how many divisions the cell as previously undergone and how many divisions the cell can undergo in the future.

Suggested biomarkers have also included resting heart rate, resting blood pressure, exercise heart rate, percent body fat, flexibility, grip strength, push strength, abdominal strength, body temperature, and skin temperature.

The present invention does not require that all of the biomarkers identified above be validated as indicative of biological age, or that they be equally useful as measures of biological age.

There is an overlap between biomarkers of aging and indicators of functional status. An indicator of functional status is an indicator that defines a functional ability (e.g., physiological, cognitive or physical function). An indicator of functional status may also be related to the increase in morbidity and mortality with chronological age. Such indicators preferably predict physiological, cognitive and physical function in an age-coherent way, and do so better than chronological age. Preferably, they can predict

the years of remaining functionality, and the trajectory toward organ-specific illness in the individual. Also, they are preferably minimally invasive.

Suggested indicators include anthropometric data (body mass index, body composition, bone density, etc.), functional challenge tests (glucose tolerance, forced vital capacity), physiological tests (cholesterol/HDL, glycosylated hemoglobin, homocysteine, etc.) and proteomic tests.

A number of mouse models for human aging exist. See Troen, supra, Table 3. The drugs identified by the present invention may be further screened in one or more of these models.

#### **Age-Related Phenotype**

An age-related phenotype is an observable change which occurs with age. An age-related phenotype may, but need not, also be a biomarker of biological aging.

Preferably, the agent of the present invention favorably affects at least one age-related phenotype. More preferably, it favorably affects at least two age-related phenotypes, more preferably phenotypes of at least two different body systems.

The age-related phenotype may be a system level phenotype, such as a measure of the condition of the nervous system, respiratory system, immune system, circulatory system, endocrine system, reproductive system, gastrointestinal system, or musculoskeletal system.

The age-related phenotype may be an organ level phenotype, such as a measure of the condition of the brain, eyes, ears, lungs, spleen, heart, pancreas, liver, ovaries, testicles, thyroid, prostate, stomach, intestines, or kidney.

The age-related phenotype may be a tissue level phenotype, such as a measure of the condition of the muscle, skin, connective tissue, nerves, or bones.

- 5 The age-related phenotype may be a cellular level phenotype, such as a measure of the condition of the cell wall, mitochondria or chromosomes.

- 10 The age-related phenotype may be a molecular level phenotype, such as a measure of the condition of nucleic acids, lipids, proteins, oxidants, and anti-oxidants.

- 15 The age-related phenotype may be manifested in a biological fluid, such as blood, urine, saliva, lymphatic fluid or cerebrospinal fluid. The biochemical composition of these fluid may be an overall, system level, organ level, tissue level, etc. phenotype, depending on the specific biochemical and fluid involved.

20

#### PHYSIOLOGICAL AGING OF THE HUMAN BODY BY SYSTEMS

SKIN, HAIR, NAILS	Loss of subcutaneous fat, Thinning of skin, Decreased collagen, Nails brittle and flake, Mucous membranes drier, Less sweat glands, Temperature regulation difficult, Hair pigment decreases, Hair thins. Eyelids baggy and wrinkled.
----------------------	--

EYES AND VISION	<p>Eyes deeper in sockets; Conjunctiva thinner and yellow; Quantity of tears decreases; Iris fades; Pupils smaller, let in less light; Night and depth vision less; "Floaters" can appear</p> <p>Lens enlarges; Lens becomes less transparent, can actually become clouded, results in cataracts; Accommodation decreases, results in presbyopia; Impaired color vision, also - especially greens and blues-- because cones degenerate; Predisposed to glaucoma (Increased pressure in eye, decreased absorption of intraocular fluid; can result in blindness);</p> <p>Macular degeneration becoming more frequent (This is the patch of retina where lens focuses light, Ultimately results in blindness )</p>
EARS AND HEARING LOSS	<p>Irreversible, sensorineural loss (presbycusis) with age (Men more affected than women, Loss occurs in higher range of sound, By 60 years, most adults have trouble hearing above 4000Hz, Normal speech 500-2000Hz)</p>
RESPIRATORY SYSTEM	<p>Lungs become more rigid, Pulmonary function decreases, Number and size of alveoli decreases, Vital capacity declines, Reduction in respiratory fluid, Bony changes in chest cavity</p>
CARDIOVASCULAR SYSTEM	<p>Heart smaller and less elastic with age, By age 70 cardiac output reduced 70%, Heart valves become sclerotic, Heart muscle more irritable, More arrhythmias, Arteries more rigid, Veins dilate</p>
GASTROINTESTINAL SYSTEM	<p>Reduced GI secretions, Reduced GI motility, Decreased weight of liver, Reduced regenerative capacity of liver, Liver metabolizes less efficiently</p>

5

10

RENAL SYSTEM	After 40 renal function decreases, By 90 lose 50% of function, Filtration and reabsorption reduced, Size and number of nephrons decrease, Bladder muscles weaken, Less able to clear drugs from system, Smaller kidneys and bladder
REPRODUCTIVE SYSTEM (MALE)	Reduced testosterone level, Testes atrophy and soften, Decrease in sperm production, Seminal fluid decreases and more viscous, Erections take more time, Refractory period after ejaculation may lengthen to days
REPRODUCTIVE SYSTEM (FEMALE)	Declining estrogen and progesterone levels, Ovulation ceases, Introitus constricts and loses elasticity, Vagina atrophies - shorter and drier, Uterus shrinks, Breasts pendulous and lose elasticity
NEUROLOGICAL SYSTEM	Neurons of central and peripheral nervous system degenerate, Nerve transmission slows, Hypothalamus less effective in regulating body temperature, Reduced REM sleep, decreased deep sleep, After age 50, lose 1% of neurons each year
MUSCULOSCELETAL SYSTEM	Adipose tissue increases with age, Lean body mass decreases, Bone mineral content diminished, Decrease in height from narrow vertebral spaces, Less resilient connective tissue, Synovial fluid more viscous, May have exaggerated curvature of spine
IMMUNE SYSTEM	Decline in immune function, Trouble differentiating between self and non-self - more auto-immune problems, Decreases antibody response, Fatty marrow replaced red marrow, Vitamin B12 absorption might decrease - decreased hemoglobin and hematocrit

5

10

ENDOCRINE SYSTEM	Decreased ability to tolerate stress - best seen in glucose metabolism, Estrogen levels decrease in women, Other hormonal decreases include testosterone, aldosterone, cortisol, progesterone
---------------------	--

Adapted from [http://www.texashte.com/html/ger\\_pap1.ppt](http://www.texashte.com/html/ger_pap1.ppt)

5

The aging human liver appears to preserve its morphology and function relatively well. The liver appears to progressively decrease in both mass and volume. It also appears browner (a condition called "brown atrophy"), as a result of accumulation of lipofuscin (ceroid) within hepatocytes. Increases occur in the number of macrohepatocytes, and in polyploidy, especially around the terminal hepatic veins. The number of mitochondria declines, and both the rough and smooth endoplasmic recticulum diminish. The number of lysozymes increase.

The liver is the premiere metabolic organ of the body. With regard to metabolism, hepatic glycerides and cholesterol levels increase with age, at least up to age 90. On the other hand, phospholipids, aminotransferases, and serum bilirubin appear to remain normal. There are contradictory reports as to the effect of aging on albumin, serum gamma-glutamyltransferase, and hepatic alkaline phosphatase. It is worth noting that it has been shown that the content of cytochrome oxidase exhibits a progressive decline which correlates with age-associated decline in mtRNA synthesis in brain, liver, heart, lungs and skeletal muscle.

See generally Anantharaju, Feller and Chedid, "Aging Liver: A Review," Gerontology, 48: 343-53 (2002).

30

#### Quality of Life

Clinicians are interested, not only in simple prolongation of lifespan, but also in maintenance of a high quality of life (QOL) over as much as possible of that lifespan. QOL can be defined subjectively in terms of the subject's

35



satisfaction with life, or objectively in terms of the subject's physical and mental ability (but not necessarily willingness) to engage in "valued activities", such as those which are pleasurable or financially rewarding.

5

Flanagan has defined five domains of QOL, capturing 15 dimensions of life quality. The five domains, and their component dimensions, are physical and material well being (Material well-being and financial security; Health and personal safety), Relations with other people (relations with spouse; Having and rearing children; Relations with parents, siblings, or other

relatives ; Relations with friends) Social, community, civic activities (Helping and encouraging others; Participating in local and governmental affairs ), Personal development, fulfillment (Intellectual development; Understanding and planning; Occupational role career; Creativity and personal expression), and recreation (Socializing with others; Passive and observational recreational activities; Participating in active recreation). See Flanagan JC, . "A research approach to improving our quality of life." Am Psychol 33:138-147 (1978).

25 "Health-related quality of life" (HRQL or HRQOL) is an individual's satisfaction or happiness with domains of life insofar as they affect or are affected by "health".

30 In a preferred embodiment, a pharmaceutical agent of the present invention is able to achieve a statistically significant improvement in the expected quality of life, measured according to a commonly accepted measure of QOL, in a treatment group over a control group.

35 While there is general acceptance of the notion that QOL is important, quantifying QOL is not especially straightforward. Also, QOL can only be measured in humans. Measurements of QOL can be objective (e.g., employment status, marital status, home ownership) or subjective (the

subject's opinion of his or her life), or some combination of the two.

A simple approach to measuring subjective QOL is to simply have the subjects rate their overall quality of life on a scale, e.g., of 7 points. One can also use more elaborate measure, such as the Older Adult Health and Mood Questionnaire (a 22 item test for assessing depression). Objective QOL can be measured by, e.g., an activities checklist.

There is a relationship between QOL assessment and so-called ADL or IADL measures, which assess the need for assistance.

The Katz Index of Independence in Activities of Daily Living (Katz ADL) measures adequacy of independent performance of bathing, dressing, toileting, transferring, continence, and feeding. See Katz, S., "Assessing Self-Maintenance: Activities of Daily Living, Mobility and Instrumental Activities of Daily Living, Journal of the American Geriatrics Society, 31(12); 721-726 (1983); Katz S., Down, T.D., Cash, H.R. et al. Progress in the Development of the Index of ADL. Gerontologist, 10:20-30 (1970).

Performance of a more sophisticated nature is measured by the "Instrumental Activities of Daily Living" (IADL) scale. This inquires into ability to independently use the telephone, shop, prepare food, carry out housekeeping, do laundry, travel locally, take medication and handle finances. See Lawton, MP and Brody, EM, Gerontologist, 9:179-86 (1969).

The 36 question Medical Outcomes Study Short Form (SF-36) (Medical Outcomes Trust, Inc., 20 Park Plaza, Suite 1014, Boston, Massachusetts 02116) assesses eight health concepts: 1) limitations in physical activities because of health problems; 2) limitations in social activities because of physical or emotional problems; 3) limitations in usual role activities because of physical health problems; 4)

bodily pain; 5) general mental health (psychological distress and well-being); 6) limitations in usual role activities because of emotional problems; 7) vitality (energy and fatigue); and 8) general health perceptions.

5

A low score on an ADL, IADL or SF-36 test is likely to be associated with a low QOL, but a high score does not guarantee a high QOL because these tests do not explore performance of "valued activities", only of more basic activities. Nonetheless, these tests can be considered commonly accepted measures of QOL for the purpose of this invention.

10

#### 15     **Age-Related Diseases**

Age-related (senescent) diseases include certain cancers, atherosclerosis, diabetes (type 2), osteoporosis, hypertension, depression, Alzheimer's, Parkinson's, glaucoma, certain immune system defects, kidney failure, and liver steatosis. In general, they are diseases for which the relative risk (comparing a subpopulation over age 55 to a suitably matched population under age 55) is at least 1.1.

20

Preferably, the agents of the present invention protect against one or more age-related diseases for at least a subpopulation of mature (post-puberty) adult subjects.

25

#### *Diabetes*

Type II diabetes is of particular interest. A deficiency of insulin in the body results in diabetes mellitus, which affects about 18 million individuals in the United States. It is characterized by a high blood glucose (sugar) level and glucose spilling into the urine due to a deficiency of insulin. As more glucose concentrates in the urine, more water is excreted, resulting in extreme thirst, rapid weight loss, drowsiness, fatigue, and possibly dehydration. Because the cells of the diabetic cannot use glucose for fuel, the body uses stored protein and fat for energy, which leads to a buildup of acid (acidosis) in the blood. If this condition is prolonged, the person can fall

30

35

into a diabetic coma, characterized by deep labored breathing and fruity-odored breath.

There are two types of diabetes mellitus, Type I and Type II. Type II diabetes is the predominant form found in the Western world; fewer than 8% of diabetic Americans have the type I disease.

*Type I diabetes.* In Type I diabetes, formerly called juvenile-onset or insulin-dependent diabetes mellitus, the pancreas cannot produce insulin. People with Type I diabetes must have daily insulin injections. But they need to avoid taking too much insulin because that can lead to insulin shock, which begins with a mild hunger. This is quickly followed by sweating, shallow breathing, dizziness, palpitations, trembling, and mental confusion. As the blood sugar falls, the body tries to compensate by breaking down fat and protein to make more sugar. Eventually, low blood sugar leads to a decrease in the sugar supply to the brain, resulting in a loss of consciousness. Eating a sugary food can prevent insulin shock until appropriate medical measures can be taken.

Type I diabetics are often characterized by their low or absent levels of circulating endogenous insulin, i.e., hypoinsulinemia (1). Islet cell antibodies causing damage to the pancreas are frequently present at diagnosis. Injection of exogenous insulin is required to prevent ketosis and sustain life.

*Type II diabetes.* Type II diabetes, formerly called adult-onset or non-insulin-dependent diabetes mellitus (NIDDM), can occur at any age. The pancreas can produce insulin, but the cells do not respond to it.

Type II diabetes is a metabolic disorder that affects approximately 17 million Americans. It is estimated that another 10 million individuals are "prone" to becoming diabetic. These vulnerable individuals can become resistant to insulin, a pancreatic hormone that signals glucose (blood sugar) uptake by fat and muscle. In order to maintain normal glucose levels, the islet cells of the pancreas produce

more insulin, resulting in a condition called hyperinsulinemia. When the pancreas can no longer produce enough insulin to compensate for the insulin resistance, and thereby maintain normal glucose levels, hyperglycemia (elevated blood glucose) results, and type II diabetes is diagnosed.

Early Type II diabetics are often characterized by hyperinsulinemia and resistance to insulin. Late Type II diabetics may be normoinsulinemic or hypoinsulinemic. Type II diabetics are usually not insulin dependent or prone to ketosis under normal circumstances.

Little is known about the disease progression from the normoinsulinemic state to the hyperinsulinemic state, and from the hyperinsulinemic state to the Type II diabetic state.

As stated above, type II diabetes is a metabolic disorder that is characterized by insulin resistance and impaired glucose-stimulated insulin secretion (2,3,4). However, Type II diabetes and atherosclerotic disease are viewed as consequences of having the insulin resistance syndrome (IRS) for many years (5). The current theory of the pathogenesis of Type II diabetes is often referred to as the "insulin resistance/islet cell exhaustion" theory. According to this theory, a condition causing insulin resistance compels the pancreatic islet cells to hypersecrete insulin in order to maintain glucose homeostasis. However, after many years of hypersecretion, the islet cells eventually fail and the symptoms of clinical diabetes are manifested. Therefore, this theory implies that, at some point, peripheral hyperinsulinemia will be an antecedent of Type II diabetes. Peripheral hyperinsulinemia can be viewed as the difference between what is produced by the beta cell minus that which is taken up by the liver. Therefore, peripheral hyperinsulinemia can be caused by increased beta cell production, decreased hepatic uptake or some combination of both. It is also important to note that it is not possible to determine the origin of insulin resistance once it is established since the onset of

peripheral hyperinsulinemia leads to a condition of global insulin resistance.

Multiple environmental and genetic factors are involved in the development of insulin resistance, hyperinsulinemia and type II diabetes. An important risk factor for the development of insulin resistance, hyperinsulinemia and type II diabetes is obesity, particularly visceral obesity (6,7,8). Type II diabetes exists world-wide, but in developed societies, the prevalence has risen as the average age of the population increases and the average individual becomes more obese.

#### **Diseases Characterized by Accelerated Aging**

Several human diseases display some features of accelerated aging. These include Werner's syndrome (classic early-onset progeria), Hutchinson-Gilford syndrome (adult progeria), and Down's syndrome (trisomy 21). Troen, Biology of Aging, Mt. Sinai J. Med., 70(1): 3 (Jan. 2003). Thus, the present invention may be useful in the treatment (curative or ameliorative) of individuals with these diseases.

#### **Genes/Proteins of Interest**

Favorable genes/proteins are those corresponding to genes less strongly expressed in longer lifespan (knockout mice) liver than in shorter lifespan (control mice) liver. Unfavorable genes/proteins are those corresponding to genes more strongly expressed in control mice liver than in knockout mice liver.

Mixed genes/proteins are those exhibiting a combination of favorable and unfavorable behavior. A mixed gene/protein can be used as would a favorable gene/protein if its favorable behavior outweighs the unfavorable. It can be used as would an unfavorable gene/protein if its unfavorable

behavior outweighs the favorable. Preferably, they are used in conjunction with other agents that affect their balance of favorable and unfavorable behavior. Use of mixed genes/proteins is, in general, less desirable than use of purely favorable or purely unfavorable genes/proteins.

For each of the differentially expressed cDNAs, corresponding mouse and human proteins have been identified, as set forth in Master Table 1. More than one human protein may be identified as corresponding to a particular mouse clone.

#### Direct and Indirect Utility of Identified Nucleic Acid Sequences and Related Molecules

The cDNAs of the disclosed clones may be used directly. For diagnostic or screening purposes, they (or specific binding fragments thereof) may be labeled and used as hybridization probes. For therapeutic purposes, they (or specific binding fragments thereof) may be used as antisense reagents to inhibit the expression of the corresponding gene, or of a sufficiently homologous gene of another species.

If the cDNA appears to be a full-length cDNA, that is, that it encodes an entire, functional protein, then it may be used in the expression of that protein. Such expression may be in cell culture, with the protein subsequently isolated and administered exogenously to subjects who would benefit therefrom, or in vivo, i.e., administration by gene therapy. Naturally, any DNA encoding the same protein, or a fragment or a mutant protein which retains the desired activity, may be used for the same purpose. The encoded protein of course has utility therapeutically and, in labeled or immobilized form, diagnostically.

The cDNAs of the disclosed clones may also be used indirectly, that is, to identify other useful DNAs, proteins, or other molecules. We have attempted to determine whether the cDNAs disclosed herein have significant similarity to any known DNA, and whether, in any

of the six possible combinations of reference frame and strand, they encode a protein similar to a known protein. If so, then it follows that the known protein, and DNAs encoding that protein, may be used in a similar manner. In addition, if the known protein is known to have additional homologues, then those homologous proteins, and DNAs encoding them, may be used in a similar manner.

There thus are several ways that a human protein homologue of interest can be identified by database searching, including but not limited to:

1) a DNA->DNA (BlastN) search for database DNAs closely related to the mouse cDNA clone identifies a particular mouse (or other nonhuman, e.g., rat) gene, and that nonhuman gene encodes a protein for which there is a known human protein homologue;

2) a DNA->Protein (BlastX) search for database proteins closely related to the translated DNA of the mouse cDNA clone identifies a particular mouse (or other nonhuman) protein, and that nonhuman protein has a known human protein homologue;

3) a DNA->DNA (BlastN) search of the database for human DNAs closely related to the mouse cDNA clone identifies a particular human DNA as a homologue of the mouse cDNA, and the corresponding human protein is known (e.g., by translation of the human DNA); and

4) a DNA->Protein (BlastX) search of the database for human proteins closely related to the translated DNA of the mouse cDNA clone identifies a particular human protein as a homologue of the corresponding mouse protein.

Thus, if we have identified a mouse cDNA, and it encodes a mouse protein which appears similar to a human protein, then that human protein may be used (especially in humans) for purposes analogous to the proposed use of the



mouse protein in mice. Moreover, a specific binding fragment of an appropriate strand of the corresponding human gene or cDNA could be labeled and used as a hybridization probe (especially against samples of human mRNA or cDNA).

5 In determining whether the disclosed cDNAs have significant similarities to known DNAs (and their translated AA sequences to known proteins), one would generally use the disclosed cDNA as a query sequence in a search of a sequence database. The results of several such searches are set  
10 forth in the Examples. Such results are dependent, to some degree, on the search parameters. Preferred parameters are set forth in Example 1. The results are also dependent on the content of the database. While the raw similarity score of a particular target (database) sequence will not vary  
15 with content (as long as it remains in the database), its informational value (in bits), expected value, and relative ranking can change. Generally speaking, the changes are small.

20 It is possible to use the sequence of the entire cDNA insert to query the database. However, the error rate increases as a sequencing run progresses. Hence, it may be beneficial to search the database using a truncated (presumably more accurate) sequence, especially if the  
25 insert is quite long.

It will be appreciated that the nucleic acid and protein databases keep growing. Hence a later search may identify high scoring target sequences which were not uncovered by an earlier search because the target sequences  
30 were not previously part of a database.

Hence, in a preferred embodiment, the cognate DNAs and proteins include not only those set forth in the examples, but those which would have been highly ranked (top ten, more preferably top three, even more preferably top two, most  
35 preferably the top one) in a search run with the same parameters on the date of filing of this application.

If the cDNA appears to be a partial cDNA, it may be used as a hybridization probe to isolate the full-length

cDNA. If the partial cDNA encodes a biologically functional fragment of the cognate protein, it may be used in a manner similar to the full length cDNA, i.e., to produce the functional fragment.

5

If we have indicated that an antagonist of a protein or other molecule is useful, then such an antagonist may be obtained by preparing a combinatorial library, as described below, of potential antagonists, and screening the library members for binding to the protein or other molecule in question. The binding members may then be further screened for the ability to antagonize the biological activity of the target. The antagonists may be used therapeutically, or, in suitably labeled or immobilized form, diagnostically.

10

15

If the cDNA is related to a known protein, then substances known to interact with that protein (e.g., agonists, antagonists, substrates, receptors, second messengers, regulators, and so forth), and binding molecules which bind them, are also of utility. Such binding molecules can likewise be identified by screening a combinatorial library.

20

#### **Isolation of Full Length cDNAs Using Partial cDNAs as probes**

If it is determined that a cDNA of the present invention is a partial cDNA, and the cognate full length cDNA is not listed in a sequence database, the available cDNA may be used as a hybridization probe to isolate the full-length cDNA from a suitable cDNA library.

25

Stringent hybridization conditions are appropriate, that is, conditions in which the hybridization temperature is 5-10 deg. C. below the  $T_m$  of the cDNA as a perfect duplex.

30

#### **Identification and Isolation of Homologous Genes/cDNAs Using a cDNA Probe**

35

It may be that the sequence databases available do not include the sequence of any homologous gene, or at least of the homologous gene for a species of interest. However,

given the cDNAs set forth above, one may readily obtain the homologous gene.

The possession of one cDNA (the "starting DNA") greatly facilitates the isolation of homologous genes/cDNAs. If the clone in question only features a partial cDNA, this partial cDNA may first be used as a probe to isolate the corresponding full length cDNA for the same species, and that the latter may be used as the starting DNA in the search for homologous genes.

The starting DNA, or a fragment thereof, is used as a hybridization probe to screen a cDNA or genomic DNA library for clones containing inserts which encode either the entire homologous protein, or a recognizable fragment thereof. The minimum length of the hybridization probe is dictated by the need for specificity. If the size of the library in bases is  $L$ , and the GC content is 50%, then the probe should have a length of at least  $l$ , where  $L = 4^l$ . This will yield, on average, a single perfect match in random DNA of  $L$  bases. The human cDNA library is about  $10^8$  bases and the human genomic DNA library is about  $10^{10}$  bases.

The library is preferably derived from an organism which is known, on biochemical evidence, to produce a homologous protein, and more preferably from the genomic DNA or mRNA of cells of that organism which are likely to be relatively high producers of that protein. A cDNA library (which is derived from an mRNA library) is especially preferred.

If the organism in question is known to have substantially different codon preferences from that of the organism whose relevant cDNA or genomic DNA is known, a synthetic hybridization probe may be used which encodes the same amino acid sequence but whose codon utilization is more similar to that of the DNA of the target organism. Alternatively, the synthetic probe may employ inosine as a substitute for those bases which are most likely to be divergent, or the probe may be a mixed probe which mixes the codons for the source DNA with the preferred codons (encoding the same amino acid) for the target organism.

By routine methods, the  $T_m$  of a perfect duplex of starting DNA is determined. One may then select a hybridization temperature which is sufficiently lower than the perfect duplex  $T_m$  to allow hybridization of the starting DNA (or other probe) to a target DNA which is divergent from the starting DNA. A 1% sequence divergence typically lowers the  $T_m$  of a duplex by 1-2°C, and the DNAs encoding homologous proteins of different species typically have sequence identities of around 50-80%. Preferably, the library is screened under conditions where the temperature is at least 20°C., more preferably at least 50°C., below the perfect duplex  $T_m$ . Since salt reduces the  $T_m$ , one ordinarily would carry out the search for DNAs encoding highly homologous proteins under relatively low salt hybridization conditions, e.g., <1M NaCl. The higher the salt concentration, and/or the lower the temperature, the greater the sequence divergence which is tolerated.

For the use of probes to identify homologous genes in other species, see, e.g., Schwinn, et al., J. Biol. Chem., 265:8183-89 (1990) (hamster 67-bp cDNA probe vs. human leukocyte genomic library; human 0.32kb DNA probe vs. bovine brain cDNA library, both with hybridization at 42°C in 6xSSC); Jenkins et al., J. Biol. Chem., 265:19624-31 (1990) (Chicken 770-bp cDNA probe vs. human genomic libraries; hybridization at 40°C in 50% formamide and 5xSSC); Murata et al., J. Exp. Med., 175:341-51 (1992) (1.2-kb mouse cDNA probe v. human eosinophil cDNA library; hybridization at 65°C in 6xSSC); Guyer et al., J. Biol. Chem., 265:17307-17 (1990) (2.95-kb human genomic DNA probe vs. porcine genomic DNA library; hybridization at 42°C in 5xSSC). The conditions set forth in these articles may each be considered suitable for the purpose of isolating homologous genes.

### Homologous Proteins and DNAs

A human protein can be said to be identifiable as homologous to a mouse cDNA clone if

(1) it can be aligned directly to the mouse cDNA clone by BlastX. and/or

(2) it can be aligned to a human gene by BlastX, whose genomic DNA (gDNA) or cDNA (DNA complementary to messenger RNA) in turn can be aligned to the mouse cDNA clone by BlastN, and/or

(3) it can be aligned to a mouse gene by BlastX, whose gDNA or cDNA in turn can be aligned to the mouse cDNA clone by BlastN, and/or

(4) it can be aligned to a mouse protein by BlastP, which in turn can be aligned to the mouse cDNA clone by BlastX, and/or

(5) it can be aligned to a mouse protein by BlastP, which in turn can be aligned to a mouse gene by BlastX, whose gDNA or cDNA can in turn be aligned to the mouse cDNA clone by BlastN;

where any alignment by BlastN, BlastP, or BlastX is in accordance with the default parameters set forth below, and the expected value (E) of each alignment (the probability that such an alignment would have occurred by chance alone) is less than  $e^{-10}$ .

A human gene is homologous to a mouse cDNA clone if it encodes a homologous human protein as defined above, or if it can be aligned either directly to the mouse cDNA clone, or indirectly through a mouse gene which can be aligned to said clone, according to the conditions set forth above. Preferably, two, three, four or all five of conditions (1)-(5) are satisfied.

Preferably, for each of conditions (1)-(5), for at least the final alignment (i.e., vs. the human protein), the E value is less than  $e^{-15}$ , more preferably less than  $e^{-20}$ , still more preferably less than  $e^{-40}$ , further more preferably less than  $e^{-50}$ , even more preferably less than  $e^{-$

60, considerably more preferably less than  $e^{-80}$ , and most preferably less than  $e^{-100}$ . More preferably, for those conditions in which the mouse cDNA clone is indirectly connected to the human protein by virtue of two or more successive alignments, the E value is so limited for all of said alignments in the connecting chain.

BlastN and BlastX report very low expected values as "0.0". This does not truly mean that the expected value is exactly zero (since any alignment could occur by chance), but merely that it is so infinitesimal that it is not reported. The documentation does not state the cutoff value, but alignments with explicit E values as low as  $e^{-178}$  (624 bits) have been reported as nonzero values, while a score of 636 bits was reported as "0.0".

Functionally homologous human proteins are also of interest. A human protein may be said to be functionally homologous to the mouse cDNA clone if (1) there is a mouse protein which is encoded by a mouse gene whose cDNA can be aligned to the mouse cDNA clone, using BlastX with the default parameters set forth below, and the E value of the alignment is less than  $e^{-50}$ , and (2) the human protein has at least one biological activity in common with the mouse protein.

The human proteins of interest also include those that are substantially and/or conservatively identical (as defined below) to the homologous and/or functionally homologous human proteins defined above.

#### Relevance of Favorable and Unfavorable Genes

If a gene is down-regulated in more favored mammals, or up-regulated in less favored mammals, (i.e., an "unfavorable gene") then several utilities are apparent.

First, the complementary strand of the gene, or a portion thereof, may be used in labeled form as a hybridization probe to detect messenger RNA and thereby monitor the level of expression of the gene in a subject. Elevated levels are

indicative of progression, or propensity to progression, to a less favored state, and clinicians may take appropriate preventative, curative or ameliorative action.

Secondly, the messenger RNA product (or equivalent cDNA), the protein product, or a binding molecule specific for that product (e.g., an antibody which binds the product), or a downstream product which mediates the activity (e.g., a signaling intermediate) or a binding molecule (e.g., an antibody) therefor, may be used, preferably in labeled or immobilized form, as an assay reagent in an assay for said nucleic acid product, protein product, or downstream product (e.g., a signaling intermediate). Again, elevated levels are indicative of a present or future problem.

Thirdly, an agent which down-regulates expression of the gene may be used to reduce levels of the corresponding protein and thereby inhibit further damage to the kidney. This agent could inhibit transcription of the gene in the subject, or translation of the corresponding messenger RNA. Possible inhibitors of transcription and translation include antisense molecules and repressor molecules. The agent could also inhibit a post-translational modification (e.g., glycosylation, phosphorylation, cleavage, GPI attachment) required for activity, or post-translationally modify the protein so as to inactivate it. Or it could be an agent which down- or up-regulated a positive or negative regulatory gene, respectively.

Fourthly, an agent which is an antagonist of the messenger RNA product or protein product of the gene, or of a downstream product through which its activity is manifested (e.g., a signaling intermediate), may be used to inhibit its activity. This antagonist could be an antibody.

Fifthly, an agent which degrades, or abets the degradation of, that messenger RNA, its protein product or a downstream product which mediates its activity (e.g., a signaling intermediate), may be used to curb the effective period of activity of the protein.

If a gene is up-regulated in more favored mammals, or down-regulated in less favored animals then the utilities are converse to those stated above.

First, the complementary strand of the gene, or a portion thereof, may be used in labeled form as a hybridization probe to detect messenger RNA and thereby monitor the level of expression of the gene in a subject. Depressed levels are indicative of damage, or possibly of a propensity to damage, and clinicians may take appropriate preventative, curative or ameliorative action.

Secondly, the messenger RNA product, the equivalent cDNA, protein product, or a binding molecule specific for those products, or a downstream product, or a signaling intermediate, or a binding molecule therefor, may be used, preferably in labeled or immobilized form, as an assay reagent in an assay for said protein product or downstream product. Again, depressed levels are indicative of a present or future problem.

Thirdly, an agent which up-regulates expression of the gene may be used to increase levels of the corresponding protein and thereby inhibit further progression to a less favored state. By way of example, it could be a vector which carries a copy of the gene, but which expresses the gene at higher levels than does the endogenous expression system. Or it could be an agent which up- or down-regulates a positive or negative regulatory gene.

Fourthly, an agent which is an agonist of the protein product of the gene, or of a downstream product through which its activity (of inhibition of progression to a less favored state) is manifested, or of a signaling intermediate may be used to foster its activity.

Fifthly, an agent which inhibits the degradation of that protein product or of a downstream product or of a signaling intermediate may be used to increase the effective period of activity of the protein.

#### Mutant Proteins



The present invention also contemplates mutant proteins (peptides) which are substantially identical (as defined below) to the parental protein (peptide). In general, the fewer the mutations, the more likely the mutant protein is to retain the activity of the parental protein. The effect of mutations is usually (but not always) additive. Certain individual mutations are more likely to be tolerated than others.

A protein is more likely to tolerate a mutation which

(a) is a substitution rather than an insertion or deletion;

(b) is an insertion or deletion at the terminus, rather than internally, or, if internal, is at a domain boundary, or a loop or turn, rather than in an alpha helix or beta strand;

(c) affects a surface residue rather than an interior residue;

(d) affects a part of the molecule distal to the binding site;

(e) is a substitution of one amino acid for another of similar size, charge, and/or hydrophobicity, and does not destroy a disulfide bond or other crosslink; and

(f) is at a site which is subject to substantial variation among a family of homologous proteins to which the protein of interest belongs.

These considerations can be used to design functional mutants.

### *Surface vs. Interior Residues*

Charged amino acid residues almost always lie on the surface of the protein. For uncharged residues, there is less certainty, but in general, hydrophilic residues are partitioned to the surface and hydrophobic residues to the interior. Of course, for a membrane protein, the membrane-spanning segments are likely to be rich in hydrophobic residues.

Surface residues may be identified experimentally by various labeling techniques, or by 3-D structure mapping

techniques like X-ray diffraction and NMR. A 3-D model of a homologous protein can be helpful.

### *Binding Site Residues*

Residues forming the binding site may be identified by (1) comparing the effects of labeling the surface residues before and after complexing the protein to its target, (2) labeling the binding site directly with affinity ligands, (3) fragmenting the protein and testing the fragments for binding activity, and (4) systematic mutagenesis (e.g., alanine-scanning mutagenesis) to determine which mutants destroy binding. If the binding site of a homologous protein is known, the binding site may be postulated by analogy.

Protein libraries may be constructed and screened that a large family (e.g.,  $10^8$ ) of related mutants may be evaluated simultaneously.

Hence, the mutations are preferably conservative modifications as defined below.

### "Substantially Identical"

A mutant protein (peptide) is substantially identical to a reference protein (peptide) if (a) it has at least 10% of a specific binding activity or a non-nutritional biological activity of the reference protein, and (b) is at least 50% identical in amino acid sequence to the reference protein (peptide). It is "substantially structurally identical" if condition (b) applies, regardless of (a).

Percentage amino acid identity is determined by aligning the mutant and reference sequences according to a rigorous dynamic programming algorithm which globally aligns their sequences to maximize their similarity, the similarity being scored as the sum of scores for each aligned pair according to an unbiased PAM250 matrix, and a penalty for each internal gap of -12 for the first null of the gap and -4 for each additional null of the same gap. The percentage identity is the number of matches expressed as a percentage of the adjusted (i.e., counting inserted nulls) length of the reference sequence.

A mutant DNA sequence is substantially identical to a reference DNA sequence if they are structural sequences, and encoding mutant and reference proteins which are substantially identical as described above.

5 If instead they are regulatory sequences, they are substantially identical if the mutant sequence has at least 10% of the regulatory activity of the reference sequence, and is at least 50% identical in nucleotide sequence to the reference sequence. Percentage identity is determined as  
10 for proteins except that matches are scored +5, mismatches -4, the gap open penalty is -12, and the gap extension penalty (per additional null) is -4.

More preferably, the sequence is not merely substantially identical but rather is at least 51%, at least  
15 66%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical in sequence to the reference sequence.

DNA sequences may also be considered "substantially identical" if they hybridize to each other under stringent conditions, i.e., conditions at which the  $T_m$  of the heteroduplex of the one strand of the mutant DNA and the more complementary strand of the reference DNA is not in excess of 10°C. less than the  $T_m$  of the reference DNA  
20 homoduplex. Typically this will correspond to a percentage  
25 identity of 85-90%.

#### "Conservative Modifications"

"Conservative modifications" are defined as

- 30 (a) conservative substitutions of amino acids as hereafter defined; or  
(b) single or multiple insertions (extension) or deletions (truncation) of amino acids at the termini.

35 Conservative modifications are preferred to other modifications. Conservative substitutions are preferred to other conservative modifications.

"Semi-Conservative Modifications" are modifications which are not conservative, but which are (a) semi-

conservative substitutions as hereafter defined; or (b) single or multiple insertions or deletions internally, but at interdomain boundaries, in loops or in other segments of relatively high mobility. Semi-conservative modifications are preferred to nonconservative modifications. Semi-conservative substitutions are preferred to other semi-conservative modifications.

Non-conservative substitutions are preferred to other non-conservative modifications.

The term "conservative" is used here in an a priori sense, i.e., modifications which would be expected to preserve 3D structure and activity, based on analysis of the naturally occurring families of homologous proteins and of past experience with the effects of deliberate mutagenesis, rather than post facto, a modification already known to conserve activity. Of course, a modification which is conservative a priori may, and usually is, also conservative post facto.

Preferably, except at the termini, no more than about five amino acids are inserted or deleted at a particular locus, and the modifications are outside regions known to contain binding sites important to activity.

Preferably, insertions or deletions are limited to the termini.

A conservative substitution is a substitution of one amino acid for another of the same exchange group, the exchange groups being defined as follows

- I Gly, Pro, Ser, Ala (Cys) (and any nonbiogenic, neutral amino acid with a hydrophobicity not exceeding that of the aforementioned a.a.'s)
- II Arg, Lys, His (and any nonbiogenic, positively-charged amino acids)
- III Asp, Glu, Asn, Gln (and any nonbiogenic negatively-charged amino acids)
- IV Leu, Ile, Met, Val (Cys) (and any nonbiogenic, aliphatic, neutral amino acid with a hydrophobicity too high for I above)

V Phe, Trp, Tyr (and any nonbiogenic, aromatic neutral amino acid with a hydrophobicity too high for I above).

Note that Cys belongs to both I and IV.

5 Residues Pro, Gly and Cys have special conformational roles. Cys participates in formation of disulfide bonds. Gly imparts flexibility to the chain. Pro imparts rigidity to the chain and disrupts  $\alpha$  helices. These residues may be essential in certain regions of the polypeptide, but  
10 substitutable elsewhere.

One, two or three conservative substitutions are more likely to be tolerated than a larger number.

"Semi-conservative substitutions" are defined herein as being substitutions within supergroup I/II/III or within  
15 supergroup IV/V, but not within a single one of groups I-V. They also include replacement of any other amino acid with alanine. If a substitution is not conservative, it preferably is semi-conservative.

"Non-conservative substitutions" are substitutions  
20 which are not "conservative" or "semi-conservative".

"Highly conservative substitutions" are a subset of conservative substitutions, and are exchanges of amino acids within the groups Phe/Tyr/Trp, Met/Leu/Ile/Val, His/Arg/Lys, Asp/Glu and Ser/Thr/Ala. They are more likely to be  
25 tolerated than other conservative substitutions. Again, the smaller the number of substitutions, the more likely they are to be tolerated.

#### "Conservatively Identical"

30 A protein (peptide) is conservatively identical to a reference protein (peptide) if it differs from the latter, if at all, solely by conservative modifications, the protein (peptide) remaining at least seven amino acids long if the reference protein (peptide) was at least seven amino acids  
35 long.

A protein is at least semi-conservatively identical to a reference protein (peptide) if it differs from the latter, if at all, solely by semi-conservative or conservative modifications.

A protein (peptide) is nearly conservatively identical to a reference protein (peptide) if it differs from the latter, if at all, solely by one or more conservative modifications and/or a single nonconservative substitution.

It is highly conservatively identical if it differs, if at all, solely by highly conservative substitutions. Highly conservatively identical proteins are preferred to those merely conservatively identical. An absolutely identical protein is even more preferred.

The core sequence of a reference protein (peptide) is the largest single fragment which retains at least 10% of a particular specific binding activity, if one is specified, or otherwise of at least one specific binding activity of the referent. If the referent has more than one specific binding activity, it may have more than one core sequence, and these may overlap or not.

If it is taught that a peptide of the present invention may have a particular similarity relationship (e.g., markedly identical) to a reference protein (peptide), preferred peptides are those which comprise a sequence having that relationship to a core sequence of the reference protein (peptide), but with internal insertions or deletions in either sequence excluded. Even more preferred peptides are those whose entire sequence has that relationship, with the same exclusion, to a core sequence of that reference protein (peptide).

### Library

The term "library" generally refers to a collection of chemical or biological entities which are related in origin, structure, and/or function, and which can be screened simultaneously for a property of interest.

Libraries may be classified by how they are constructed (natural vs. artificial diversity; combinatorial vs. noncombinatorial), how they are screened (hybridization,

expression, display), or by the nature of the screened library members (peptides, nucleic acids, etc.).

In a "natural diversity" library, essentially all of the diversity arose without human intervention. This would be true, for example, of messenger RNA extracted from a non-engineered cell.

In a "synthetic diversity" library, essentially all of the diversity arose deliberately as a result of human intervention. This would be true for example of a combinatorial library; note that a small level of natural diversity could still arise as a result of spontaneous mutation. It would also be true of a noncombinatorial library of compounds collected from diverse sources, even if they were all natural products.

In a "non-natural diversity" library, at least some of the diversity arose deliberately through human intervention.

In a "controlled origin" library, the source of the diversity is limited in some way. A limitation might be to cells of a particular individual, to a particular species, or to a particular genus, or, more complexly, to individuals of a particular species who are of a particular age, sex, physical condition, geographical location, occupation and/or familial relationship. Alternatively or additionally, it might be to cells of a particular tissue or organ. Or it could be cells exposed to particular pharmacological, environmental, or pathogenic conditions. Or the library could be of chemicals, or a particular class of chemicals, produced by such cells.

In a "controlled structure" library, the library members are deliberately limited by the production conditions to particular chemical structures. For example, if they are oligomers, they may be limited in length and monomer composition, e.g. hexapeptides composed of the twenty genetically encoded amino acids.

#### Hybridization Library

In a hybridization library, the library members are nucleic acids, and are screened using a nucleic acid

hybridization probe. Bound nucleic acids may then be amplified, cloned, and/or sequenced.

### Expression Library

5           In an expression library, the screened library members are gene expression products, but one may also speak of an underlying library of genes encoding those products. The library is made by subcloning DNA encoding the library members (or portions thereof) into expression vectors (or  
10 into cloning vectors which subsequently are used to construct expression vectors), each vector comprising an expressible gene encoding a particular library member, introducing the expression vectors into suitable cells, and expressing the genes so the expression products are  
15 produced.

          In one embodiment, the expression products are secreted, so the library can be screened using an affinity reagent, such as an antibody or receptor. The bound expression products may be sequenced directly, or their  
20 sequences inferred by, e.g., sequencing at least the variable portion of the encoding DNA.

          In a second embodiment, the cells are lysed, thereby exposing the expression products, and the latter are screened with the affinity reagent.

25           In a third embodiment, the cells express the library members in such a manner that they are displayed on the surface of the cells, or on the surface of viral particles produced by the cells. (See display libraries, below).

          In a fourth embodiment, the screening is not for the  
30 ability of the expression product to bind to an affinity reagent, but rather for its ability to alter the phenotype of the host cell in a particular detectable manner. Here, the screened library members are transformed cells, but there is a first underlying library of expression products  
35 which mediate the behavior of the cells, and a second underlying library of genes which encode those products.

### Display Library



In a display library, the library members are each conjugated to, and displayed upon, a support of some kind. The support may be living (a cell or virus), or nonliving (e.g., a bead or plate).

5        If the support is a cell or virus, display will normally be effectuated by expressing a fusion protein which comprises the library member, a carrier moiety allowing integration of the fusion protein into the surface of the cell or virus, and optionally a lining moiety. In a  
10       variation on this theme, the cell coexpresses a first fusion comprising the library member and a linking moiety L1, and a second fusion comprising a linking moiety L2 and the carrier moiety. L1 and L2 interact to associate the first fusion with the second fusion and hence, indirectly, the library  
15       member with the surface of the cell or virus.

#### Soluble Library

In a soluble library, the library members are free in solution. A soluble library may be produced directly, or  
20       one may first make a display library and then release the library members from their supports.

#### Encapsulated Library

In an encapsulated library, the library members are  
25       inside cells or liposomes. Generally speaking, encapsulated libraries are used to store the library members for future use; the members are extracted in some way for screening purposes. However, if they differentially affect the phenotype of the cells, they may be screened indirectly by  
30       screening the cells.

#### cDNA Library

A cDNA library is usually prepared by extracting RNA from cells of particular origin, fractionating the RNA to  
35       isolate the messenger RNA (mRNA has a poly(A) tail, so this is usually done by oligo-dT affinity chromatography), synthesizing complementary DNA (cDNA) using reverse transcriptase, DNA polymerase, and other enzymes, subcloning the cDNA into vectors, and introducing the vectors into

cells. Often, only mRNAs or cDNAs of particular sizes will be used, to make it more likely that the cDNA encodes a functional polypeptide.

A cDNA library explores the natural diversity of the transcribed DNAs of cells from a particular source. It is not a combinatorial library.

A cDNA library may be used to make a hybridization library, or it may be used as an (or to make) expression library.

### Genomic DNA Library

A genomic DNA library is made by extracting DNA from a particular source, fragmenting the DNA, isolating fragments of a particular size range, subcloning the DNA fragments into vectors, and introducing the vectors into cells.

Like a cDNA library, a genomic DNA library is a natural diversity library, and not a combinatorial library. A genomic DNA library may be used the same way as a cDNA library.

### Synthetic DNA library

A synthetic DNA library may be screened directly (as a hybridization library), or used in the creation of an expression or display library of peptides/proteins.

### Combinatorial Libraries

The term "combinatorial library" refers to a library in which the individual members are either systematic or random combinations of a limited set of basic elements, the properties of each member being dependent on the choice and location of the elements incorporated into it. Typically, the members of the library are at least capable of being screened simultaneously. Randomization may be complete or partial; some positions may be randomized and others predetermined, and at random positions, the choices may be limited in a predetermined manner. The members of a combinatorial library may be oligomers or polymers of some kind, in which the variation occurs through the choice of monomeric building block at one or more positions of the

oligomer or polymer, and possibly in terms of the connecting linkage, or the length of the oligomer or polymer, too. Or the members may be nonoligomeric molecules with a standard core structure, like the 1,4-benzodiazepine structure, with the variation being introduced by the choice of substituents at particular variable sites on the core structure. Or the members may be nonoligomeric molecules assembled like a jigsaw puzzle, but wherein each piece has both one or more variable moieties (contributing to library diversity) and one or more constant moieties (providing the functionalities for coupling the piece in question to other pieces).

Thus, in a typical combinatorial library, chemical building blocks are at least partially randomly combined into a large number (as high as  $10^{15}$ ) of different compounds, which are then simultaneously screened for binding (or other) activity against one or more targets.

In a "simple combinatorial library", all of the members belong to the same class of compounds (e.g., peptides) and can be synthesized simultaneously. A "composite combinatorial library" is a mixture of two or more simple libraries, e.g., DNAs and peptides, or peptides, peptoids, and PNAs, or benzodiazepines and carbamates. The number of component simple libraries in a composite library will, of course, normally be smaller than the average number of members in each simple library, as otherwise the advantage of a library over individual synthesis is small.

Libraries of thousands, even millions, of random oligopeptides have been prepared by chemical synthesis (Houghten et al., *Nature*, 354:84-6(1991)), or gene expression (Marks et al., *J Mol Biol*, 222:581-97(1991)), displayed on chromatographic supports (Lam et al., *Nature*, 354:82-4(1991)), inside bacterial cells (Colas et al., *Nature*, 380:548-550(1996)), on bacterial pili (Lu, *Bio/Technology*, 13:366-372(1990)), or phage (Smith, *Science*, 228:1315-7(1985)), and screened for binding to a variety of targets including antibodies (Valadon et al., *J Mol Biol*, 261:11-22(1996)), cellular proteins (Schmitz et al., *J Mol Biol*, 260:664-677(1996)), viral proteins (Hong and Boulanger, *Embo J*, 14:4714-4727(1995)), bacterial proteins

(Jacobsson and Frykberg, *Biotechniques*, 18:878-885(1995)), nucleic acids (Cheng et al., *Gene*, 171:1-8(1996)), and plastic (Siani et al., *J Chem Inf Comput Sci*, 34:588-593(1994)).

5 Libraries of proteins (Ladner, USP 4,664,989), peptoids (Simon et al., *Proc Natl Acad Sci U S A*, 89:9367-71(1992)), nucleic acids (Ellington and Szostak, *Nature*, 246:818(1990)), carbohydrates, and small organic molecules (Eichler et al., *Med Res Rev*, 15:481-96(1995)) have also  
10 been prepared or suggested for drug screening purposes.

The first combinatorial libraries were composed of peptides or proteins, in which all or selected amino acid positions were randomized. Peptides and proteins can exhibit high and specific binding activity, and can act as  
15 catalysts. In consequence, they are of great importance in biological systems.

Nucleic acids have also been used in combinatorial libraries. Their great advantage is the ease with which a nucleic acid with appropriate binding activity can be  
20 amplified. As a result, combinatorial libraries composed of nucleic acids can be of low redundancy and hence, of high diversity.

There has also been much interest in combinatorial libraries based on small molecules, which are more suited to  
25 pharmaceutical use, especially those which, like benzodiazepines, belong to a chemical class which has already yielded useful pharmacological agents. The techniques of combinatorial chemistry have been recognized as the most efficient means for finding small molecules that  
30 act on these targets. At present, small molecule combinatorial chemistry involves the synthesis of either pooled or discrete molecules that present varying arrays of functionality on a common scaffold. These compounds are grouped in libraries that are then screened against the  
35 target of interest either for binding or for inhibition of biological activity.

The size of a library is the number of molecules in it. The simple diversity of a library is the number of unique structures in it. There is no formal minimum or maximum

diversity. If the library has a very low diversity, the library has little advantage over just synthesizing and screening the members individually. If the library is of very high diversity, it may be inconvenient to handle, at least without automatizing the process. The simple diversity of a library is preferably at least  $10^1$ ,  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^6$ ,  $10^7$ ,  $10^8$  or  $10^9$ , the higher the better under most circumstances. The simple diversity is usually not more than  $10^{15}$ , and more usually not more than  $10^{10}$ . The average sampling level is the size divided by the simple diversity. The expected average sampling level must be high enough to provide a reasonable assurance that, if a given structure were expected, as a consequence of the library design, to be present, that the actual average sampling level will be high enough so that the structure, if satisfying the screening criteria, will yield a positive result when the library is screened. Thus, the preferred average sampling level is a function of the detection limit, which in turn is a function of the strength of the signal to be screened.

There are more complex measures of diversity than simple diversity. These attempt to take into account the degree of structural difference between the various unique sequences. These more complex measures are usually used in the context of small organic compound libraries, see below.

The library members may be presented as solutes in solution, or immobilized on some form of support. In the latter case, the support may be living (cell, virus) or nonliving (bead, plate, etc.). The supports may be separable (cells, virus particles, beads) so that binding and nonbinding members can be separated, or nonseparable (plate). In the latter case, the members will normally be placed on addressable positions on the support. The advantage of a soluble library is that there is no carrier moiety that could interfere with the binding of the members to the support. The advantage of an immobilized library is that it is easier to identify the structure of the members which were positive.

When screening a soluble library, or one with a separable support, the target is usually immobilized. When screening

a library on a nonseparable support, the target will usually be labeled.

### Oligonucleotide Libraries

5       An oligonucleotide library is a combinatorial library, at least some of whose members are single-stranded oligonucleotides having three or more nucleotides connected by phosphodiester or analogous bonds. The oligonucleotides may be linear, cyclic or branched, and may include non-  
10       nucleic acid moieties. The nucleotides are not limited to the nucleotides normally found in DNA or RNA. For examples of nucleotides modified to increase nuclease resistance and chemical stability of aptamers, see Chart 1 in Osborne and Ellington, Chem. Rev., 97: 349-70 (1997). For screening of  
15       RNA, see Ellington and Szostak, Nature, 346: 818-22 (1990).

      There is no formal minimum or maximum size for these oligonucleotides. However, the number of conformations which an oligonucleotide can assume increases exponentially with its length in bases. Hence, a longer oligonucleotide is  
20       more likely to be able to fold to adapt itself to a protein surface. On the other hand, while very long molecules can be synthesized and screened, unless they provide a much superior affinity to that of shorter molecules, they are not likely to be found in the selected population, for the  
25       reasons explained by Osborne and Ellington (1997). Hence, the libraries of the present invention are preferably composed of oligonucleotides having a length of 3 to 100 bases, more preferably 15 to 35 bases. The oligonucleotides in a given library may be of the same or of different  
30       lengths.

      Oligonucleotide libraries have the advantage that libraries of very high diversity (e.g.,  $10^{15}$ ) are feasible, and binding molecules are readily amplified in vitro by polymerase chain reaction (PCR). Moreover, nucleic acid  
35       molecules can have very high specificity and affinity to targets.

      In a preferred embodiment, this invention prepares and screens oligonucleotide libraries by the SELEX method, as described in King and Famulok, Molec. Biol. Repts., 20: 97-

107 (1994); L. Gold, C. Tuerk. Methods of producing nucleic acid ligands, US#5595877; Oliphant et al. Gene 44:177 (1986).

The term "aptamer" is conferred on those oligonucleotides which bind the target protein. Such aptamers may be used to characterize the target protein, both directly (through identification of the aptamer and the points of contact between the aptamer and the protein) and indirectly (by use of the aptamer as a ligand to modify the chemical reactivity of the protein).

In a classic oligonucleotide, each nucleotide (monomeric unit) is composed of a phosphate group, a sugar moiety, and either a purine or a pyrimidine base. In DNA, the sugar is deoxyribose and in RNA it is ribose. The nucleotides are linked by 5'-3' phosphodiester bonds.

The deoxyribose phosphate backbone of DNA can be modified to increase resistance to nuclease and to increase penetration of cell membranes. Derivatives such as mono- or dithiophosphates, methyl phosphonates, boranophosphates, formacetals, carbamates, siloxanes, and dimethylenethio-sulfoxideo- and-sulfono- linked species are known in the art.

#### Peptide Library

A peptide is composed of a plurality of amino acid residues joined together by peptidyl (-NHCO-) bonds. A biogenic peptide is a peptide in which the residues are all genetically encoded amino acid residues; it is not necessary that the biogenic peptide actually be produced by gene expression.

Amino acids are the basic building blocks with which peptides and proteins are constructed. Amino acids possess both an amino group (-NH<sub>2</sub>) and a carboxylic acid group (-COOH). Many amino acids, but not all, have the alpha amino acid structure NH<sub>2</sub>-CHR-COOH, where R is hydrogen, or any of a variety of functional groups.

Twenty amino acids are genetically encoded: Alanine, Arginine, Asparagine, Aspartic Acid, Cysteine, Glutamic Acid, Glutamine, Glycine, Histidine, Isoleucine, Leucine,

Lysine, Methionine, Phenylalanine, Proline, Serine, Threonine, Tryptophan, Tyrosine, and Valine. Of these, all save Glycine are optically isomeric, however, only the L-form is found in humans. Nevertheless, the D-forms of these amino acids do have biological significance; D-Phe, for example, is a known analgesic.

Many other amino acids are also known, including: 2-Aminoadipic acid; 3-Aminoadipic acid; beta-Aminopropionic acid; 2-Aminobutyric acid; 4-Aminobutyric acid (Piperidinic acid); 6-Aminocaproic acid; 2-Aminoheptanoic acid; 2-Aminoisobutyric acid, 3-Aminoisobutyric acid; 2-Aminopimelic acid; 2,4-Diaminobutyric acid; Desmosine; 2,2'-Diaminopimelic acid; 2,3-Diaminopropionic acid; N-Ethylglycine; N-Ethylasparagine; Hydroxylysine; allo-Hydroxylysine; 3-Hydroxyproline; 4-Hydroxyproline; Isodesmosine; allo-Isoleucine; N-Methylglycine (Sarcosine); N-Methylisoleucine; N-Methylvaline; Norvaline; Norleucine; and Ornithine.

Peptides are constructed by condensation of amino acids and/or smaller peptides. The amino group of one amino acid (or peptide) reacts with the carboxylic acid group of a second amino acid (or peptide) to form a peptide (-NHCO-) bond, releasing one molecule of water. Therefore, when an amino acid is incorporated into a peptide, it should, technically speaking, be referred to as an amino acid residue. The core of that residue is the moiety which excludes the -NH and -CO linking functionalities which connect it to other residues. This moiety consists of one or more main chain atoms (see below) and the attached side chains.

The main chain moiety of each amino acid consists of the -NH and -CO linking functionalities and a core main chain moiety. Usually the latter is a single carbon atom. However, the core main chain moiety may include additional carbon atoms, and may also include nitrogen, oxygen or sulfur atoms, which together form a single chain. In a preferred embodiment, the core main chain atoms consist solely of carbon atoms.



The side chains are attached to the core main chain atoms. For alpha amino acids, in which the side chain is attached to the alpha carbon, the C-1, C-2 and N-2 of each residue form the repeating unit of the main chain, and the word "side chain" refers to the C-3 and higher numbered carbon atoms and their substituents. It also includes H atoms attached to the main chain atoms.

Amino acids may be classified according to the number of carbon atoms which appear in the main chain between the carbonyl carbon and amino nitrogen atoms which participate in the peptide bonds. Among the 150 or so amino acids which occur in nature, alpha, beta, gamma and delta amino acids are known. These have 1-4 intermediary carbons. Only alpha amino acids occur in proteins. Proline is a special case of an alpha amino acid; its side chain also binds to the peptide bond nitrogen.

For beta and higher order amino acids, there is a choice as to which main chain core carbon a side chain other than H is attached to. The preferred attachment site is the C-2 (alpha) carbon, i.e., the one adjacent to the carboxyl carbon of the -CO linking functionality. It is also possible for more than one main chain atom to carry a side chain other than H. However, in a preferred embodiment, only one main chain core atom carries a side chain other than H.

A main chain carbon atom may carry either one or two side chains; one is more common. A side chain may be attached to a main chain carbon atom by a single or a double bond; the former is more common.

A simple combinatorial peptide library is one whose members are peptides having three or more amino acids connected via peptide bonds.

The peptides may be linear, branched, or cyclic, and may covalently or noncovalently include nonpeptidyl moieties. The amino acids are not limited to the naturally occurring or to the genetically encoded amino acids.

A biased peptide library is one in which one or more (but not all) residues of the peptides are constant residues.

### *Cyclic Peptides*

Many naturally occurring peptides are cyclic. Cyclization is a common mechanism for stabilization of peptide conformation thereby achieving improved association of the peptide with its ligand and hence improved biological activity. Cyclization is usually achieved by intra-chain cystine formation, by formation of peptide bond between side chains or between N- and C- terminals. Cyclization was usually achieved by peptides in solution, but several publications have appeared that describe cyclization of peptides on beads.

A peptide library may be an oligopeptide library or a protein library.

### *Oligopeptides*

Preferably, the oligopeptides are at least five, six, seven or eight amino acids in length. Preferably, they are composed of less than 50, more preferably less than 20 amino acids.

In the case of an oligopeptide library, all or just some of the residues may be variable. The oligopeptide may be unconstrained, or constrained to a particular conformation by, e.g., the participation of constant cysteine residues in the formation of a constraining disulfide bond.

### *Proteins*

Proteins, like oligopeptides, are composed of a plurality of amino acids, but the term protein is usually reserved for longer peptides, which are able to fold into a stable conformation. A protein may be composed of two or more polypeptide chains, held together by covalent or noncovalent crosslinks. These may occur in a homooligomeric or a heterooligomeric state.

A peptide is considered a protein if it (1) is at least 50 amino acids long, or (2) has at least two stabilizing covalent crosslinks (e.g., disulfide bonds). Thus, conotoxins are considered proteins.

Usually, the proteins of a protein library will be characterizable as having both constant residues (the same for all proteins in the library) and variable residues (which vary from member to member). This is simply because, for a given range of variation at each position, the sequence space (simple diversity) grows exponentially with the number of residue positions, so at some point it becomes inconvenient for all residues of a peptide to be variable positions. Since proteins are usually larger than oligopeptides, it is more common for protein libraries than oligopeptide libraries to feature variable positions.

In the case of a protein library, it is desirable to focus the mutations at those sites which are tolerant of mutation. These may be determined by alanine scanning mutagenesis or by comparison of the protein sequence to that of homologous proteins of similar activity. It is also more likely that mutation of surface residues will directly affect binding. Surface residues may be determined by inspecting a 3D structure of the protein, or by labeling the surface and then ascertaining which residues have received labels. They may also be inferred by identifying regions of high hydrophilicity within the protein.

Because proteins are often altered at some sites but not others, protein libraries can be considered a special case of the biased peptide library.

There are several reasons that one might screen a protein library instead of an oligopeptide library, including (1) a particular protein, mutated in the library, has the desired activity to some degree already, and (2) the oligopeptides are not expected to have a sufficiently high affinity or specificity since they do not have a stable conformation.

When the protein library is based on a parental protein which does not have the desired activity, the parental protein will usually be one which is of high stability (melting point  $\geq 50$  deg. C.) and/or possessed of hypervariable regions.

The variable domains of an antibody possess hypervariable regions and hence, in some embodiments, the

protein library comprises members which comprise a mutant of VH or VL chain, or a mutant of an antigen-specific binding fragment of such a chain. VH and VL chains are usually each about 110 amino acid residues, and are held in proximity by a disulfide bond between the adjoining CL and CH1 regions to form a variable domain. Together, the VH, VL, CL and CH1 form an Fab fragment.

In human heavy chains, the hypervariable regions are at 31-35, 49-65, 98-111 and 84-88, but only the first three are involved in antigen binding. There is variation among VH and VL chains at residues outside the hypervariable regions, but to a much lesser degree.

A sequence is considered a mutant of a VH or VL chain if it is at least 80% identical to a naturally occurring VH or VL chain at all residues outside the hypervariable region.

In a preferred embodiment, such antibody library members comprise both at least one VH chain and at least one VL chain, at least one of which is a mutant chain, and which chains may be derived from the same or different antibodies. The VH and VL chains may be covalently joined by a suitable linker moiety, as in a "single chain antibody", or they may be noncovalently joined, as in a naturally occurring variable domain.

If the joining is noncovalent, and the library is displayed on cells or virus, then either the VH or the VL chain may be fused to the carrier surface/coat protein. The complementary chain may be co-expressed, or added exogenously to the library.

The members may further comprise some or all of an antibody constant heavy and/or constant light chain, or a mutant thereof.

#### Peptoid Library

A peptoid is an analogue of a peptide in which one or more of the peptide bonds (-NH-CO-) are replaced by pseudopeptide bonds, which may be the same or different. It is not necessary that all of the peptide bonds be replaced,

i.e., a peptoid may include one or more conventional amino acid residues, e.g., proline.

A peptide bond has two small divalent linker elements, -NH- and -CO-. Thus, a preferred class of pseudopeptide bonds are those which consist of two small divalent linker elements. Each may be chosen independently from the group consisting of amine (-NH-), substituted amine (-NR-), carbonyl (-CO-), thiocarbonyl (-CS-), methylene (-CH<sub>2</sub>-), monosubstituted methylene (-CHR-), disubstituted methylene (-CR<sub>1</sub>R<sub>2</sub>-), ether (-O-) and thioether (-S-). The more preferred pseudopeptide bonds include:

N-modified -NRCO-

Carba Ψ -CH<sub>2</sub>-CH<sub>2</sub>-

Depsi Ψ -CO-O-

Hydroxyethylene Ψ -CHOH-CH<sub>2</sub>-

Ketomethylene Ψ -CO-CH<sub>2</sub>-

Methylene-Oxy -CH<sub>2</sub>-O-

Reduced -CH<sub>2</sub>-NH-

Thiomethylene -CH<sub>2</sub>-S-

Thiopeptide -CS-NH-

Retro-Inverso -CO-NH-

A single peptoid molecule may include more than one kind of pseudopeptide bond.

For the purposes of introducing diversity into a peptoid library, one may vary (1) the side chains attached to the core main chain atoms of the monomers linked by the pseudopeptide bonds, and/or (2) the side chains (e.g., the R of an -NRCO-) of the pseudopeptide bonds. Thus, in one embodiment, the monomeric units which are not amino acid residues are of the structure -NR<sub>1</sub>-CR<sub>2</sub>-CO-, where at least one of R<sub>1</sub> and R<sub>2</sub> are not hydrogen. If there is variability in the pseudopeptide bond, this is most conveniently done by using an -NRCO- or other pseudopeptide bond with an R group, and varying the R group. In this event, the R group will usually be any of the side chains characterizing the amino acids of peptides, as previously discussed.

If the R group of the pseudopeptide bond is not variable, it will usually be small, e.g., not more than 10

atoms (e.g., hydroxyl, amino, carboxyl, methyl, ethyl, propyl).

If the conjugation chemistries are compatible, a simple combinatorial library may include both peptides and peptoids.

#### Peptide Nucleic Acid Library

A PNA oligomer is here defined as one comprising a plurality of units, at least one of which is a PNA monomer which comprises a side chain comprising a nucleobase. For nucleobases, see USP 6,077,835.

The classic PNA oligomer is composed of (2-aminoethyl)glycine units, with nucleobases attached by methylene carbonyl linkers. That is, it has the structure



where the outer parenthesized substructure is the PNA monomer.

In this structure, the nucleobase B is separated from the backbone N by three bonds, and the points of attachment of the side chains are separated by six bonds. The nucleobase may be any of the bases included in the nucleotides discussed in connection with oligonucleotide libraries. The bases of nucleotides A, G, T, C and U are preferred.

A PNA oligomer may further comprise one or more amino acid residues, especially glycine and proline.

One can readily envision related molecules in which (1) the  $-\text{COCH}_2-$  linker is replaced by another linker, especially one composed of two small divalent linkers as defined previously, (2) a side chain is attached to one of the three main chain carbons not participating in the peptide bond (either instead or in addition to the side chain attached to the N of the classic PNA); and/or (3) the peptide bonds are replaced by pseudopeptide bonds as disclosed previously in the context of peptoids.

PNA oligomer libraries have been made; see e.g. Cook, 6,204,326.

### Small Organic Compound Library

5       The small organic compound library ("compound library", for short) is a combinatorial library whose members are suitable for use as drugs if, indeed, they have the ability to mediate a biological activity of the target protein.

10       Peptides have certain disadvantages as drugs. These include susceptibility to degradation by serum proteases, and difficulty in penetrating cell membranes. Preferably, all or most of the compounds of the compound library avoid, or at least do not suffer to the same degree, one or more of the pharmaceutical disadvantages of peptides.

15       In designing a compound library, it is helpful to bear in mind the methods of molecular modification typically used to obtain new drugs. Three basic kinds of modification may be identified: disjunction, in which a lead drug is simplified to identify its component pharmacophoric  
20       moieties; conjunction, in which two or more known pharmacophoric moieties, which may be the same or different, are associated, covalently or noncovalently, to form a new drug; and alteration, in which one moiety is replaced by another which may be similar or different, but which is not  
25       in effect a disjunction or conjunction. The use of the terms "disjunction", "conjunction" and "alteration" is intended only to connote the structural relationship of the end product to the original leads, and not how the new drugs are actually synthesized, although it is possible that the  
30       two are the same.

      The process of disjunction is illustrated by the evolution of neostigmine (1931) and edrophonium (1952) from physostigmine (1925). Subsequent conjunction is illustrated by demecarium (1956) and ambenonium (1956).

35       Alterations may modify the size, polarity, or electron distribution of an original moiety. Alterations include ring closing or opening, formation of lower or higher homologues, introduction or saturation of double bonds, introduction of optically active centers, introduction,

removal or replacement of bulky groups, isosteric or bioisosteric substitution, changes in the position or orientation of a group, introduction of alkylating groups, and introduction, removal or replacement of groups with a view toward inhibiting or promoting inductive (electrostatic) or conjugative (resonance) effects.

Thus, the substituents may include electron acceptors and/or electron donors. Typical electron donors (+I) include  $-\text{CH}_3$ ,  $-\text{CH}_2\text{R}$ ,  $-\text{CHR}_2$ ,  $-\text{CR}_3$  and  $-\text{COO}^-$ . Typical electron acceptors (-I) include  $-\text{NH}_3^+$ ,  $-\text{NR}_3^+$ ,  $-\text{NO}_2$ ,  $-\text{CN}$ ,  $-\text{COOH}$ ,  $-\text{COOR}$ ,  $-\text{CHO}$ ,  $-\text{COR}$ ,  $-\text{COR}$ ,  $-\text{F}$ ,  $-\text{Cl}$ ,  $-\text{Br}$ ,  $-\text{OH}$ ,  $-\text{OR}$ ,  $-\text{SH}$ ,  $-\text{SR}$ ,  $-\text{CH}=\text{CH}_2$ ,  $-\text{CR}=\text{CR}_2$ , and  $-\text{C}=\text{CH}$ .

The substituents may also include those which increase or decrease electronic density in conjugated systems. The former (+R) groups include  $-\text{CH}_3$ ,  $-\text{CR}_3$ ,  $-\text{F}$ ,  $-\text{Cl}$ ,  $-\text{Br}$ ,  $-\text{I}$ ,  $-\text{OH}$ ,  $-\text{OR}$ ,  $-\text{OCOR}$ ,  $-\text{SH}$ ,  $-\text{SR}$ ,  $-\text{NH}_2$ ,  $-\text{NR}_2$ , and  $-\text{NHCOR}$ . The later (-R) groups include  $-\text{NO}_2$ ,  $-\text{CN}$ ,  $-\text{CHC}$ ,  $-\text{COR}$ ,  $-\text{COOH}$ ,  $-\text{COOR}$ ,  $-\text{CONH}_2$ ,  $-\text{SO}_2\text{R}$  and  $-\text{CF}_3$ .

Synthetically speaking, the modifications may be achieved by a variety of unit processes, including nucleophilic and electrophilic substitution, reduction and oxidation, addition elimination, double bond cleavage, and cyclization.

For the purpose of constructing a library, a compound, or a family of compounds, having one or more pharmacological activities (which need not be related to the known or suspected activities of the target protein), may be disjoined into two or more known or potential pharmacophoric moieties. Analogues of each of these moieties may be identified, and mixtures of these analogues reacted so as to reassemble compounds which have some similarity to the original lead compound. It is not necessary that all members of the library possess moieties analogous to all of the moieties of the lead compound.

The design of a library may be illustrated by the example of the benzodiazepines. Several benzodiazepine drugs, including chlordiazepoxide, diazepam and oxazepam, have been used as anti-anxiety drugs. Derivatives of benzodiazepines have widespread biological activities;



derivatives have been reported to act not only as anxiolytics, but also as anticonvulsants; cholecystokinin (CCK) receptor subtype A or B, kappa opioid receptor, platelet activating factor, and HIV transactivator Tat antagonists, and GPIIbIIIa, reverse transcriptase and ras farnesyltransferase inhibitors.

The benzodiazepine structure has been disjoined into a 2-aminobenzophenone, an amino acid, and an alkylating agent. See Bunin, et al., Proc. Nat. Acad. Sci. USA, 91:4708 (1994). Since only a few 2-aminobenzophenone derivatives are commercially available, it was later disjoined into 2-aminoarylstannane, an acid chloride, an amino acid, and an alkylating agent. Bunin, et al., Meth. Enzymol., 267:448 (1996). The arylstannane may be considered the core structure upon which the other moieties are substituted, or all four may be considered equals which are conjoined to make each library member.

A basic library synthesis plan and member structure is shown in Figure 1 of Fowlkes, et al., U.S. Serial No. 08/740,671, incorporated by reference in its entirety. The acid chloride building block introduces variability at the R<sup>1</sup> site. The R<sup>2</sup> site is introduced by the amino acid, and the R<sup>3</sup> site by the alkylating agent. The R<sup>4</sup> site is inherent in the arylstannane. Bunin, et al. generated a 1, 4-benzodiazepine library of 11,200 different derivatives prepared from 20 acid chlorides, 35 amino acids, and 16 alkylating agents. (No diversity was introduced at R<sup>4</sup>; this group was used to couple the molecule to a solid phase.) According to the Available Chemicals Directory (HDL Information Systems, San Leandro CA), over 300 acid chlorides, 80 Fmoc-protected amino acids and 800 alkylating agents were available for purchase (and more, of course, could be synthesized). The particular moieties used were chosen to maximize structural dispersion, while limiting the numbers to those conveniently synthesized in the wells of a microtiter plate. In choosing between structurally similar compounds, preference was given to the least substituted compound.

The variable elements included both aliphatic and aromatic groups. Among the aliphatic groups, both acyclic and cyclic (mono- or poly-) structures, substituted or not, were tested. (While all of the acyclic groups were linear, it would have been feasible to introduce a branched aliphatic). The aromatic groups featured either single and multiple rings, fused or not, substituted or not, and with heteroatoms or not. The secondary substituents included -NH<sub>2</sub>, -OH, -OMe, -CN, -Cl, -F, and -COOH. While not used, spacer moieties, such as -O-, -S-, -OO-, -CS-, -NH-, and -NR-, could have been incorporated.

Bunin et al. suggest that instead of using a 1, 4-benzodiazepine as a core structure, one may instead use a 1, 4-benzodiazepine-2, 5-dione structure.

As noted by Bunin et al., it is advantageous, although not necessary, to use a linkage strategy which leaves no trace of the linking functionality, as this permits construction of a more diverse library.

Other combinatorial nonoligomeric compound libraries known or suggested in the art have been based on carbamates, mercaptoacylated pyrrolidines, phenolic agents, aminimides, N-acylamino ethers (made from amino alcohols, aromatic hydroxy acids, and carboxylic acids), N-alkylamino ethers (made from aromatic hydroxy acids, amino alcohols and aldehydes) 1, 4-piperazines, and 1, 4-piperazine-6-ones.

DeWitt, et al., Proc. Nat. Acad. Sci. (USA), 90:6909-13 (1993) describe the simultaneous but separate, synthesis of 40 discrete hydantoins and 40 discrete benzodiazepines. They carry out their synthesis on a solid support (inside a gas dispersion tube), in an array format, as opposed to other conventional simultaneous synthesis techniques (e.g., in a well, or on a pin). The hydantoins were synthesized by first simultaneously deprotecting and then treating each of five amino acid resins with each of eight isocyanates. The benzodiazepines were synthesized by treating each of five deprotected amino acid resins with each of eight 2-amino benzophenone imines.

Chen, et al., J. Am. Chem. Soc., 116:2661-62 (1994) described the preparation of a pilot (9 member)

combinatorial library of formate esters. A polymer bead-bound aldehyde preparation was "split" into three aliquots, each reacted with one of three different ylide reagents. The reaction products were combined, and then divided into three new aliquots, each of which was reacted with a different Michael donor. Compound identity was found to be determinable on a single bead basis by gas chromatography/mass spectroscopy analysis.

Holmes, USP 5,549,974 (1996) sets forth methodologies for the combinatorial synthesis of libraries of thiazolidinones and metathiazanones. These libraries are made by combination of amines, carbonyl compounds, and thiols under cyclization conditions.

Ellman, USP 5,545,568 (1996) describes combinatorial synthesis of benzodiazepines, prostaglandins, beta-turn mimetics, and glycerol-based compounds. See also Ellman, USP 5,288,514.

Summerton, USP 5,506,337 (1996) discloses methods of preparing a combinatorial library formed predominantly of morpholino subunit structures.

Heterocyclic combinatorial libraries are reviewed generally in Nefzi, et al., Chem. Rev., 97:449-472 (1997).

For pharmacological classes, see, e.g., Goth, Medical Pharmacology: Principles and Concepts (C.V. Mosby Co.: 8th ed. 1976); Korolkovas and Burckhalter, Essentials of Medicinal Chemistry (John Wiley & Sons, Inc.: 1976). For synthetic methods, see, e.g., Warren, Organic Synthesis: The Disconnection Approach (John Wiley & Sons, Ltd.: 1982); Fuson, Reactions of Organic Compounds (John Wiley & Sons: 1966); Payne and Payne, How to do an Organic Synthesis (Allyn and Bacon, Inc.: 1969); Greene, Protective Groups in Organic Synthesis (Wiley-Interscience). For selection of substituents, see e.g., Hansch and Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology (John Wiley & Sons: 1979).

The library is preferably synthesized so that the individual members remain identifiable so that, if a member is shown to be active, it is not necessary to analyze it.

Several methods of identification have been proposed, including:

- (1) encoding, i.e., the attachment to each member of an identifier moiety which is more readily identified than the member proper. This has the disadvantage that the tag may itself influence the activity of the conjugate.
- (2) spatial addressing, e.g., each member is synthesized only at a particular coordinate on or in a matrix, or in a particular chamber. This might be, for example, the location of a particular pin, or a particular well on a microtiter plate, or inside a "tea bag".

The present invention is not limited to any particular form of identification.

However, it is possible to simply characterize those members of the library which are found to be active, based on the characteristic spectroscopic indicia of the various building blocks.

Solid phase synthesis permits greater control over which derivatives are formed. However, the solid phase could interfere with activity. To overcome this problem, some or all of the molecules of each member could be liberated, after synthesis but before screening.

Examples of candidate simple libraries which might be evaluated include derivatives of the following:

Cyclic Compounds Containing One Hetero Atom

Heteronitrogen

pyrroles

pentasubstituted pyrroles

pyrrolidines

pyrrolines

prolines

indoles

beta-carbolines

pyridines

dihydropyridines

1,4-dihydropyridines

pyrido[2,3-d]pyrimidines

tetrahydro-3H-imidazo[4,5-c] pyridines

Isoquinolines

tetrahydroisoquinolines

quinolones

5                    beta-lactams  
                    azabicyclo[4.3.0]nonen-8-one amino acid

Heterooxygen

furans

tetrahydrofurans

10                    2,5-disubstituted tetrahydrofurans

pyrans

hydroxypyranones

tetrahydroxypyranones

gamma-butyrolactones

15                    Heterosulfur

sulfolenes

Cyclic Compounds with Two or More Hetero atoms

Multiple heteronitrogens

imidazoles

20                    pyrazoles

piperazines

diketopiperazines

arylpiperazines

benzylpiperazines

25                    benzodiazepines

1,4-benzodiazepine-2,5-diones

hydantoins

5-alkoxyhydantoins

dihydropyrimidines

30                    1,3-disubstituted-5,6-dihydropyrimidine-2,4-diones

cyclic ureas

cyclic thioureas

35                    quinazolines

chiral 3-substituted-quinazoline-2,4-

diones

triazoles

1,2,3-triazoles

purines

Heteronitrogen and Heterooxygen

dikelomorpholines

isoxazoles

5

isoxazolines

Heteronitrogen and Heterosulfur

thiazolidines

N-axylthiazolidines

dihydrothiazoles

10

2-methylene-2,3-dihydrothiazates

2-aminothiazoles

thiophenes

3-amino thiophenes

4-thiazolidinones

15

4-melathiazanones

benzisothiazolones

For details on synthesis of libraries, see Nefzi, et al., Chem. Rev., 97:449-72 (1997), and references cited therein.

20

#### **Pharmaceutical Methods and Preparations**

The preferred animal subject of the present invention is a mammal. By the term "mammal" is meant an individual belonging to the class Mammalia. The invention is particularly useful in the treatment of human subjects, although it is intended for veterinary and nutritional uses as well. Preferred nonhuman subjects are of the orders Primata (e.g., apes and monkeys), Artiodactyla or Perissodactyla (e.g., cows, pigs, sheep, horses, goats), Carnivora (e.g., cats, dogs), Rodenta (e.g., rats, mice, guinea pigs, hamsters), Lagomorpha (e.g., rabbits) or other pet, farm or laboratory mammals.

30

The term "protection", as used herein, is intended to include "prevention," "suppression" and "treatment."

35

"Prevention", strictly speaking, involves administration of the pharmaceutical prior to the induction of the disease (or other adverse clinical condition). "Suppression" involves administration of the composition prior to the clinical appearance of the disease. "Treatment" involves

administration of the protective composition after the appearance of the disease.

It will be understood that in human and veterinary medicine, it is not always possible to distinguish between "preventing" and "suppressing" since the ultimate inductive event or events may be unknown, latent, or the patient is not ascertained until well after the occurrence of the event or events. Therefore, unless qualified, the term "prevention" will be understood to refer to both prevention in the strict sense, and to suppression.

The preventative or prophylactic use of a pharmaceutical involves identifying subjects who are at higher risk than the general population of contracting the disease, and administering the pharmaceutical to them in advance of the clinical appearance of the disease. The effectiveness of such use is measured by comparing the subsequent incidence or severity of the disease, or of particular symptoms of the disease, in the treated subjects against that in untreated subjects of the same high risk group.

While high risk factors vary from disease to disease, in general, these include (1) prior occurrence of the disease in one or more members of the same family, or, in the case of a contagious disease, in individuals with whom the subject has come into potentially contagious contact at a time when the earlier victim was likely to be contagious, (2) a prior occurrence of the disease in the subject, (3) prior occurrence of a related disease, or a condition known to increase the likelihood of the disease, in the subject; (4) appearance of a suspicious level of a marker of the disease, or a related disease or condition; (5) a subject who is immunologically compromised, e.g., by radiation treatment, HIV infection, drug use, etc., or (6) membership in a particular group (e.g., a particular age, sex, race, ethnic group, etc.) which has been epidemiologically associated with that disease.

A prophylaxis or treatment may be curative, that is, directed at the underlying cause of a disease, or

ameliorative, that is, directed at the symptoms of the disease, especially those which reduce the quality of life.

It should also be understood that to be useful, the protection provided need not be absolute, provided that it is sufficient to carry clinical value. An agent which provides protection to a lesser degree than do competitive agents may still be of value if the other agents are ineffective for a particular individual, if it can be used in combination with other agents to enhance the level of protection, or if it is safer than competitive agents. It is desirable that there be a statistically significant ( $p=0.05$  or less) improvement in the treated subject relative to an appropriate untreated control, and it is desirable that this improvement be at least 10%, more preferably at least 25%, still more preferably at least 50%, even more preferably at least 100%, in some indicia of the incidence or severity of the disease or of at least one symptom of the disease.

At least one of the drugs of the present invention may be administered, by any means that achieve their intended purpose, to protect a subject against a disease or other adverse condition. The form of administration may be systemic or topical. For example, administration of such a composition may be by various parenteral routes such as subcutaneous, intravenous, intradermal, intramuscular, intraperitoneal, intranasal, transdermal, or buccal routes. Alternatively, or concurrently, administration may be by the oral route. Parenteral administration can be by bolus injection or by gradual perfusion over time.

A typical regimen comprises administration of an effective amount of the drug, administered over a period ranging from a single dose, to dosing over a period of hours, days, weeks, months, or years.

It is understood that the suitable dosage of a drug of the present invention will be dependent upon the age, sex, health, and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment, and the nature of the effect desired. However, the most preferred dosage can be tailored to the individual subject, as is understood and determinable by one of skill in the art, without undue



experimentation. This will typically involve adjustment of a standard dose, e.g., reduction of the dose if the patient has a low body weight.

Prior to use in humans, a drug will first be evaluated for safety and efficacy in laboratory animals. In human clinical studies, one would begin with a dose expected to be safe in humans, based on the preclinical data for the drug in question, and on customary doses for analogous drugs (if any). If this dose is effective, the dosage may be decreased, to determine the minimum effective dose, if desired. If this dose is ineffective, it will be cautiously increased, with the patients monitored for signs of side effects. See, e.g., Berkow et al, eds., *The Merck Manual*, 15th edition, Merck and Co., Rahway, N.J., 1987; Goodman et al., eds., *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 8th edition, Pergamon Press, Inc., Elmsford, N.Y., (1990); Avery's *Drug Treatment: Principles and Practice of Clinical Pharmacology and Therapeutics*, 3rd edition, ADIS Press, LTD., Williams and Wilkins, Baltimore, MD. (1987), Ebadi, *Pharmacology*, Little, Brown and Co., Boston, (1985), which references and references cited therein, are entirely incorporated herein by reference.

The total dose required for each treatment may be administered by multiple doses or in a single dose. The protein may be administered alone or in conjunction with other therapeutics directed to the disease or directed to other symptoms thereof.

The appropriate dosage form will depend on the disease, the pharmaceutical, and the mode of administration; possibilities include tablets, capsules, lozenges, dental pastes, suppositories, inhalants, solutions, ointments and parenteral depots. See, e.g., Berker, *supra*, Goodman, *supra*, Avery, *supra* and Ebadi, *supra*, which are entirely incorporated herein by reference, including all references cited therein.

In the case of peptide drugs, the drug may be administered in the form of an expression vector comprising a nucleic acid encoding the peptide; such a vector, after incorporation into the genetic complement of a cell of the

patient, directs synthesis of the peptide. Suitable vectors include genetically engineered poxviruses (vaccinia), adenoviruses, adeno-associated viruses, herpesviruses and lentiviruses which are or have been rendered nonpathogenic.

5 In addition to at least one drug as described herein, a pharmaceutical composition may contain suitable pharmaceutically acceptable carriers, such as excipients, carriers and/or auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. See, e.g., Berker, *supra*, Goodman, *supra*, 10 Avery, *supra* and Ebadi, *supra*, which are entirely incorporated herein by reference, included all references cited therein.

## 15 **Assay Compositions and Methods**

### Target Organism

The invention contemplates that it may be appropriate to ascertain or to mediate the biological activity of a substance of this invention in a target organism.

20 The target organism may be a plant, animal, or microorganism.

In the case of a plant, it may be an economic plant, in which case the drug may be intended to increase the disease, weather or pest resistance, alter the growth 25 characteristics, or otherwise improve the useful characteristics or mute undesirable characteristics of the plant. Or it may be a weed, in which case the drug may be intended to kill or otherwise inhibit the growth of the plant, or to alter its characteristics to convert it from a weed to an economic plant. The plant may be a tree, shrub, 30 crop, grass, etc. The plant may be an algae (which are in some cases also microorganisms), or a vascular plant, especially gymnosperms (particularly conifers) and angiosperms. Angiosperms may be monocots or dicots. The plants of greatest interest are rice, wheat, corn, alfalfa, 35 soybeans, potatoes, peanuts, tomatoes, melons, apples, pears, plums, pineapples, fir, spruce, pine, cedar, and oak.

If the target organism is a microorganism, it may be algae, bacteria, fungi, or a virus (although the biological

activity of a virus must be determined in a virus-infected cell). The microorganism may be human or other animal or plant pathogen, or it may be nonpathogenic. It may be a soil or water organism, or one which normally lives inside other living things.

If the target organism is an animal, it may be a vertebrate or a nonvertebrate animal. Nonvertebrate animals are chiefly of interest when they act as pathogens or parasites, and the drugs are intended to act as biocidic or biostatic agents. Nonvertebrate animals of interest include worms, mollusks, and arthropods.

The target organism may also be a vertebrate animal, i.e., a mammal, bird, reptile, fish or amphibian. Among mammals, the target animal preferably belongs to the order Primata (humans, apes and monkeys), Artiodactyla (e.g., cows, pigs, sheep, goats, horses), Rodenta (e.g., mice, rats) Lagomorpha (e.g., rabbits, hares), or Carnivora (e.g., cats, dogs). Among birds, the target animals are preferably of the orders Anseriformes (e.g., ducks, geese, swans) or Galliformes (e.g., quails, grouse, pheasants, turkeys and chickens). Among fish, the target animal is preferably of the order Clupeiformes (e.g., sardines, shad, anchovies, whitefish, salmon).

#### Target Tissues

The term "target tissue" refers to any whole animal, physiological system, whole organ, part of organ, miscellaneous tissue, cell, or cell component (e.g., the cell membrane) of a target animal in which biological activity may be measured.

Routinely in mammals one would choose to compare and contrast the biological impact on virtually any and all tissues which express the subject receptor protein. The main tissues to use are: brain, heart, lung, kidney, liver, pancreas, skin, intestines, adipose, stomach, skeletal muscle, adrenal glands, breast, prostate, vasculature, retina, cornea, thyroid gland, parathyroid glands, thymus, bone marrow, bone, etc.

Another classification would be by cell type: B cells, T cells, macrophages, neutrophils, eosinophils, mast cells, platelets, megakaryocytes, erythrocytes, bone marrow stromal cells, fibroblasts, neurons, astrocytes, neuroglia, microglia, epithelial cells (from any organ, e.g. skin, breast, prostate, lung, intestines etc), cardiac muscle cells, smooth muscle cells, striated muscle cells, osteoblasts, osteocytes, chondroblasts, chondrocytes, keratinocytes, melanocytes, etc.

Of course, in the case of a unicellular organism, there is no distinction between the "target organism" and the "target tissue".

### Screening Assays

Assays intended to determine the binding or the biological activity of a substance are called preliminary screening assays.

Screening assays will typically be either in vitro (cell-free) assays (for binding to an immobilized receptor) or cell-based assays (for alterations in the phenotype of the cell). They will not involve screening of whole multicellular organisms, or isolated organs. The comments on diagnostic biological assays apply mutatis mutandis to screening cell-based assays.

### In Vitro vs. In Vivo Assays

The term *in vivo* is descriptive of an event, such as binding or enzymatic action, which occurs within a living organism. The organism in question may, however, be genetically modified. The term *in vitro* refers to an event which occurs outside a living organism. Parts of an organism (e.g., a membrane, or an isolated biochemical) are used, together with artificial substrates and/or conditions. For the purpose of the present invention, the term *in vitro* excludes events occurring inside or on an intact cell, whether of a unicellular or multicellular organism.

*In vivo* assays include both cell-based assays, and organismic assays. The cell-based assays include both assays on unicellular organisms, and assays on isolated cells or

cell cultures derived from multicellular organisms. The cell cultures may be mixed, provided that they are not organized into tissues or organs. The term organismic assay refers to assays on whole multicellular organisms, and assays on isolated organs or tissues of such organisms.

### In vitro Diagnostic Methods and Reagents

The in vitro assays of the present invention may be applied to any suitable analyte-containing sample, and may be qualitative or quantitative in nature.

#### *Sample*

The sample will normally be a biological fluid, such as blood, urine, lymph, semen, milk, or cerebrospinal fluid, or a fraction or derivative thereof, or a biological tissue, in the form of, e.g., a tissue section or homogenate. However, the sample conceivably could be (or derived from) a food or beverage, a pharmaceutical or diagnostic composition, soil, or surface or ground water. If a biological fluid or tissue, it may be taken from a human or other mammal, vertebrate or animal, or from a plant. The preferred sample is blood, or a fraction or derivative thereof.

#### *Binding and Reaction Assays*

The assay may be a binding assay, in which one step involves the binding of a diagnostic reagent to the analyte, or a reaction assay, which involves the reaction of a reagent with the analyte. The reagents used in a binding assay may be classified as to the nature of their interaction with analyte: (1) analyte analogues, or (2) analyte binding molecules (ABM). They may be labeled or insolubilized.

In a reaction assay, the assay may look for a direct reaction between the analyte and a reagent which is reactive with the analyte, or if the analyte is an enzyme or enzyme inhibitor, for a reaction catalyzed or inhibited by the analyte. The reagent may be a reactant, a catalyst, or an inhibitor for the reaction.

An assay may involve a cascade of steps in which the product of one step acts as the target for the next step. These steps may be binding steps, reaction steps, or a combination thereof.

5

#### *Signal Producing System (SPS)*

In order to detect the presence, or measure the amount, of an analyte, the assay must provide for a signal producing system (SPS) in which there is a detectable difference in the signal produced, depending on whether the analyte is present or absent (or, in a quantitative assay, on the amount of the analyte). The detectable signal may be one which is visually detectable, or one detectable only with instruments. Possible signals include production of colored or luminescent products, alteration of the characteristics (including amplitude or polarization) of absorption or emission of radiation by an assay component or product, and precipitation or agglutination of a component or product. The term "signal" is intended to include the discontinuance of an existing signal, or a change in the rate of change of an observable parameter, rather than a change in its absolute value. The signal may be monitored manually or automatically.

In a reaction assay, the signal is often a product of the reaction. In a binding assay, it is normally provided by a label borne by a labeled reagent.

#### *Labels*

The component of the signal producing system which is most intimately associated with the diagnostic reagent is called the "label". A label may be, e.g., a radioisotope, a fluorophore, an enzyme, a co-enzyme, an enzyme substrate, an electron-dense compound, an agglutinable particle.

The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography. Isotopes which are particularly useful for the purpose of the present invention include  $^3\text{H}$ ,  $^{125}\text{I}$ ,

$^{131}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ ,  $^{32}\text{P}$  and  $^{33}\text{P}$ .  $^{125}\text{I}$  is preferred for antibody labeling.

The label may also be a fluorophore. When the fluorescently labeled reagent is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labelling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

Alternatively, fluorescence-emitting metals such as  $^{125}\text{Eu}$ , or others of the lanthanide series, may be incorporated into a diagnostic reagent using such metal chelating groups as diethylenetriaminepentaacetic acid (DTPA) or ethylenediamine-tetraacetic acid (EDTA).

The label may also be a chemiluminescent compound. The presence of the chemiluminescently labeled reagent is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isolumino, thermotropic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used for labeling. Bioluminescence is a type of chemiluminescence found in biological systems in which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

Enzyme labels, such as horseradish peroxidase and alkaline phosphatase, are preferred. When an enzyme label is used, the signal producing system must also include a substrate for the enzyme. If the enzymatic reaction product is not itself detectable, the SPS will include one or more additional reactants so that a detectable product appears.

An enzyme analyte may act as its own label if an enzyme inhibitor is used as a diagnostic reagent.

Binding assays may be divided into two basic types, heterogeneous and homogeneous. In heterogeneous assays, the interaction between the affinity molecule and the analyte does not affect the label, hence, to determine the amount or presence of analyte, bound label must be separated from free label. In homogeneous assays, the interaction does affect the activity of the label, and therefore analyte levels can be deduced without the need for a separation step.

In one embodiment, the ABM is insolubilized by coupling it to a macromolecular support, and analyte in the sample is allowed to compete with a known quantity of a labeled or specifically labelable analyte analogue. The "analyte analogue" is a molecule capable of competing with analyte for binding to the ABM, and the term is intended to include analyte itself. It may be labeled already, or it may be labeled subsequently by specifically binding the label to a moiety differentiating the analyte analogue from analyte. The solid and liquid phases are separated, and the labeled analyte analogue in one phase is quantified. The higher the level of analyte analogue in the solid phase, i.e., sticking to the ABM, the lower the level of analyte in the sample.

In a "sandwich assay", both an insolubilized ABM, and a labeled ABM are employed. The analyte is captured by the insolubilized ABM and is tagged by the labeled ABM, forming a ternary complex. The reagents may be added to the sample in either order, or simultaneously. The ABMs may be the same or different. The amount of labeled ABM in the ternary complex is directly proportional to the amount of analyte in the sample.

The two embodiments described above are both heterogeneous assays. However, homogeneous assays are conceivable. The key is that the label be affected by whether or not the complex is formed.

#### Conjugation Methods

A label may be conjugated, directly or indirectly (e.g., through a labeled anti-ABM antibody), covalently (e.g., with SPDP) or noncovalently, to the ABM, to produce a diagnostic reagent. Similarly, the ABM may be conjugated to



a solid phase support to form a solid phase ("capture") diagnostic reagent.

Suitable supports include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, agaroses, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention.

The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to its target. Thus the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc.

#### Biological Assays

A biological assay measures or detects a biological response of a biological entity to a substance.

The biological entity may be a whole organism, an isolated organ or tissue, freshly isolated cells, an immortalized cell line, or a subcellular component (such as a membrane; this term should not be construed as including an isolated receptor). The entity may be, or may be derived from, an organism which occurs in nature, or which is modified in some way. Modifications may be genetic (including radiation and chemical mutants, and genetic engineering) or somatic (e.g., surgical, chemical, etc.). In the case of a multicellular entity, the modifications may affect some or all cells. The entity need not be the target organism, or a derivative thereof, if there is a reasonable correlation between bioassay activity in the assay entity and biological activity in the target organism.

The entity is placed in a particular environment, which may be more or less natural. For example, a culture medium may, but need not, contain serum or serum substitutes, and it may, but need not, include a support matrix of some kind, it may be still, or agitated. It may contain particular biological or chemical agents, or have particular physical

parameters (e.g., temperature), that are intended to nourish or challenge the biological entity.

There must also be a detectable biological marker for the response. At the cellular level, the most common markers are cell survival and proliferation, cell behavior (clustering, motility), cell morphology (shape, color), and biochemical activity (overall DNA synthesis, overall protein synthesis, and specific metabolic activities, such as utilization of particular nutrients, e.g., consumption of oxygen, production of CO<sub>2</sub>, production of organic acids, uptake or discharge of ions).

The direct signal produced by the biological marker may be transformed by a signal producing system into a different signal which is more observable, for example, a fluorescent or colorimetric signal.

The entity, environment, marker and signal producing system are chosen to achieve a clinically acceptable level of sensitivity, specificity and accuracy.

In some cases, the goal will be to identify substances which mediate the biological activity of a natural biological entity, and the assay is carried out directly with that entity. In other cases, the biological entity is used simply as a model of some more complex (or otherwise inconvenient to work with) biological entity. In that event, the model biological entity is used because activity in the model system is considered more predictive of activity in the ultimate natural biological entity than is simple binding activity in an in vitro system. The model entity is used instead of the ultimate entity because the former is more expensive or slower to work with, or because ethical considerations forbid working with the ultimate entity yet.

The model entity may be naturally occurring, if the model entity usefully models the ultimate entity under some conditions. Or it may be non-naturally occurring, with modifications that increase its resemblance to the ultimate entity.

Transgenic animals, such as transgenic mice, rats, and rabbits, have been found useful as model systems.

In cell-based model assays, where the biological activity is mediated by binding to a receptor (target protein), the receptor may be functionally connected to a signal (biological marker) producing system, which may be endogenous or exogenous to the cell.

There are a number of techniques of doing this.

#### "Zero-Hybrid" Systems

In these systems, the binding of a peptide to the target protein results in a screenable or selectable phenotypic change, without resort to fusing the target protein (or a ligand binding moiety thereof) to an endogenous protein. It may be that the target protein is endogenous to the host cell, or is substantially identical to an endogenous receptor so that it can take advantage of the latter's native signal transduction pathway. Or sufficient elements of the signal transduction pathway normally associated with the target protein may be engineered into the cell so that the cell signals binding to the target protein.

#### "One-Hybrid" Systems

In these systems, a chimera receptor, a hybrid of the target protein and an endogenous receptor, is used. The chimeric receptor has the ligand binding characteristics of the target protein and the signal transduction characteristics of the endogenous receptor. Thus, the normal signal transduction pathway of the endogenous receptor is subverted.

Preferably, the endogenous receptor is inactivated, or the conditions of the assay avoid activation of the endogenous receptor, to improve the signal-to-noise ratio.

See Fowlkes USP 5,789,184 for a yeast system.

Another type of "one-hybrid" system combines a peptide: DNA-binding domain fusion with an unfused target receptor that possesses an activation domain.

#### "Two-Hybrid" System

In a preferred embodiment, the cell-based assay is a two hybrid system. This term implies that the ligand is incorporated into a first hybrid protein, and the receptor into a second hybrid protein. The first hybrid also  
5 comprises component A of a signal generating system, and the second hybrid comprises component B of that system. Components A and B, by themselves, are insufficient to generate a signal. However, if the ligand binds the receptor, components A and B are brought into sufficiently  
10 close proximity so that they can cooperate to generate a signal.

Components A and B may naturally occur, or be substantially identical to moieties which naturally occur, as components of a single naturally occurring biomolecule,  
15 or they may naturally occur, or be substantially identical to moieties which naturally occur, as separate naturally occurring biomolecules which interact in nature.

#### Two-Hybrid System: Transcription Factor Type

20 In a preferred "two-hybrid" embodiment, one member of a peptide ligand:receptor binding pair is expressed as a fusion to a DNA-binding domain (DBD) from a transcription factor (this fusion protein is called the "bait"), and the other is expressed as a fusion to a transactivation domain  
25 (TAD) (this fusion protein is called the "fish", the "prey", or the "catch"). The transactivation domain should be complementary to the DNA-binding domain, i.e., it should interact with the latter so as to activate transcription of a specially designed reporter gene that carries a binding  
30 site for the DNA-binding domain. Naturally, the two fusion proteins must likewise be complementary.

This complementarity may be achieved by use of the complementary and separable DNA-binding and transcriptional activator domains of a single transcriptional activator  
35 protein, or one may use complementary domains derived from different proteins. The domains may be identical to the native domains, or mutants thereof. The assay members may be fused directly to the DBD or TAD, or fused through an intermediated linker.

The target DNA operator may be the native operator sequence, or a mutant operator. Mutations in the operator may be coordinated with mutations in the DBD and the TAD. An example of a suitable transcription activation system is one comprising the DNA-binding domain from the bacterial repressor LexA and the activation domain from the yeast transcription factor Gal4, with the reporter gene operably linked to the LexA operator.

It is not necessary to employ the intact target receptor; just the ligand-binding moiety is sufficient.

The two fusion proteins may be expressed from the same or different vectors. Likewise, the activatable reporter gene may be expressed from the same vector as either fusion protein (or both proteins), or from a third vector.

Potential DNA-binding domains include Gal4, LexA, and mutant domains substantially identical to the above.

Potential activation domains include E. coli B42, Gal4 activation domain II, and HSV VP16, and mutant domains substantially identical to the above.

Potential operators include the native operators for the desired activation domain, and mutant domains substantially identical to the native operator.

The fusion proteins may comprise nuclear localization signals.

The assay system will include a signal producing system, too. The first element of this system is a reporter gene operably linked to an operator responsive to the DBD and TAD of choice. The expression of this reporter gene will result, directly or indirectly, in a selectable or screenable phenotype (the signal). The signal producing system may include, besides the reporter gene, additional genetic or biochemical elements which cooperate in the production of the signal. Such an element could be, for example, a selective agent in the cell growth medium. There may be more than one signal producing system, and the system may include more than one reporter gene.

The sensitivity of the system may be adjusted by, e.g., use of competitive inhibitors of any step in the activation or signal production process, increasing or decreasing the

number of operators, using a stronger or weaker DBD or TAD, etc.

When the signal is the death or survival of the cell in question, or proliferation or nonproliferation of the cell in question, the assay is said to be a selection. When the signal merely results in a detectable phenotype by which the signaling cell may be differentiated from the same cell in a nonsignaling state (either way being a living cell), the assay is a screen. However, the term "screening assay" may be used in a broader sense to include a selection. When the narrower sense is intended, we will use the term "nonselective screen".

Various screening and selection systems are discussed in Ladner, USP 5,198,346.

Screening and selection may be for or against the peptide: target protein or compound:target protein interaction.

Preferred assay cells are microbial (bacterial, yeast, algal, protozoal), invertebrate, vertebrate (esp. mammalian, particularly human). The best developed two-hybrid assays are yeast and mammalian systems.

Normally, two hybrid assays are used to determine whether a protein X and a protein Y interact, by virtue of their ability to reconstitute the interaction of the DBD and the TAD. However, augmented two-hybrid assays have been used to detect interactions that depend on a third, non-protein ligand.

For more guidance on two-hybrid assays, see Brent and Finley, Jr., *Ann. Rev. Genet.*, 31:663-704 (1997); Fremont-Racine, et al., *Nature Genetics*, 277-281 (16 July 1997); Allen, et al., *TIBS*, 511-16 (Dec. 1995); LeCrenier, et al., *BioEssays*, 20:1-6 (1998); Xu, et al., *Proc. Nat. Acad. sci. (USA)*, 94:12473-8 (Nov. 1992); Esotak, et al., *Mol. Cell. Biol.*, 15:5820-9 (1995); Yang, et al., *Nucleic Acids Res.*, 23:1152-6 (1995); Bendixen, et al., *Nucleic Acids Res.*, 22:1778-9 (1994); Fuller, et al., *BioTechniques*, 25:85-92 (July 1998); Cohen, et al., *PNAS (USA)* 95:14272-7 (1998); Kolonin and Finley, Jr., *PNAS (USA)* 95:14266-71 (1998). See also Vasavada, et al., *PNAS (USA)*, 88:10686-90 (1991)

(contingent replication assay), and Rehrauer, et al., J. Biol. Chem., 271:23865-73 91996) (LexA repressor cleavage assay).

## 5 Two-Hybrid Systems: reporter Enzyme type

In another embodiment, the components A and B reconstitute an enzyme which is not a transcription factor.

10 As in the last example, the effect of the reconstitution of the enzyme is a phenotypic change which may be a screenable change, a selectable change, or both.

## In vivo Diagnostic Uses

15 Radio-labeled ABM may be administered to the human or animal subject. Administration is typically by injection, e.g., intravenous or arterial or other means of administration in a quantity sufficient to permit subsequent dynamic and/or static imaging using suitable radio-detecting devices. The dosage is the smallest amount capable of  
20 providing a diagnostically effective image, and may be determined by means conventional in the art, using known radio-imaging agents as a guide.

Typically, the imaging is carried out on the whole body of the subject, or on that portion of the body or organ  
25 relevant to the condition or disease under study. The amount of radio-labeled ABM accumulated at a given point in time in relevant target organs can then be quantified.

A particularly suitable radio-detecting device is a scintillation camera, such as a gamma camera. A  
30 scintillation camera is a stationary device that can be used to image distribution of radio-labeled ABM. The detection device in the camera senses the radioactive decay, the distribution of which can be recorded. Data produced by the imaging system can be digitized. The digitized information  
35 can be analyzed over time discontinuously or continuously. The digitized data can be processed to produce images, called frames, of the pattern of uptake of the radio-labelled ABM in the target organ at a discrete point in time. In most continuous (dynamic) studies, quantitative

data is obtained by observing changes in distributions of radioactive decay in target organs over time. In other words, a time-activity analysis of the data will illustrate uptake through clearance of the radio-labeled binding protein by the target organs with time.

Various factors should be taken into consideration in selecting an appropriate radioisotope. The radioisotope must be selected with a view to obtaining good quality resolution upon imaging, should be safe for diagnostic use in humans and animals, and should preferably have a short physical half-life so as to decrease the amount of radiation received by the body. The radioisotope used should preferably be pharmacologically inert, and, in the quantities administered, should not have any substantial physiological effect.

The ABM may be radio-labeled with different isotopes of iodine, for example  $^{123}\text{I}$ ,  $^{125}\text{I}$ , or  $^{131}\text{I}$  (see for example, U.S. Patent 4,609,725). The extent of radio-labeling must, however be monitored, since it will affect the calculations made based on the imaging results (i.e. a diiodinated ABM will result in twice the radiation count of a similar monoiodinated ABM over the same time frame).

In applications to human subjects, it may be desirable to use radioisotopes other than  $^{125}\text{I}$  for labeling in order to decrease the total dosimetry exposure of the human body and to optimize the detectability of the labeled molecule (though this radioisotope can be used if circumstances require). Ready availability for clinical use is also a factor. Accordingly, for human applications, preferred radio-labels are for example,  $^{99\text{m}}\text{Tc}$ ,  $^{67}\text{Ga}$ ,  $^{68}\text{Ga}$ ,  $^{90}\text{Y}$ ,  $^{111}\text{In}$ ,  $^{113\text{m}}\text{In}$ ,  $^{123}\text{I}$ ,  $^{186}\text{Re}$ ,  $^{188}\text{Re}$  or  $^{211}\text{At}$ .

The radio-labelled ABM may be prepared by various methods. These include radio-halogenation by the chloramine - T method or the lactoperoxidase method and subsequent purification by HPLC (high pressure liquid chromatography), for example as described by J. Gutkowska et al in "Endocrinology and Metabolism Clinics of America: (1987) 16 (1):183. Other known methods of radio-labeling can be used, such as IODOBEADS™.



There are a number of different methods of delivering the radio-labeled ABM to the end-user. It may be administered by any means that enables the active agent to reach the agent's site of action in the body of a mammal.

5 Because proteins are subject to being digested when administered orally, parenteral administration, i.e., intravenous, subcutaneous, intramuscular, would ordinarily be used to optimize absorption of an ABM, such as an antibody, which is a protein.

**EXAMPLES**

By disrupting the GHR/BP gene we are able to investigate the role GH signaling plays in a variety of disease states and in various stages of growth, development and aging. The lack of GH signaling in these mice produces a phenotype which is dwarf, long-lived, insulin sensitive, protected from type 1 diabetic kidney damage and resistant to high-fat diet induced hyperglycemia.

Another model of altered GH signaling, the dwarf GHA mouse, does not have an altered lifespan but does show protection from diabetic-induced kidney damage and resistance to diet-induced hyperglycemia (List et al., 2001; Coschigano et al., 2002; Coschigano et al., 2003). Conversely, overexpression of bGH in mice significantly shortens lifespan (Doi et al., 1988).

The objective of this study was to examine gene expression differences between livers of GHR/BP -/- and +/+ mice in an attempt to account for some of the physiological changes and to identify cDNAs that are regulated by GH signaling.

**Animal Models and Methods**

**Animal Models** 60 day old male growth hormone receptor/binding protein gene disrupted (GHR-BP -/-) mice and their respective controls in a BalbC/129 Ola genetic background were sacrificed and total liver RNA was isolated.

Table A. Summary of physiological alterations of GHR/BP -/- mice.

Physiological Parameter	% of +/+
Body Weight	52%
Body Length	76%
Liver Weight	42%
Plasma GH Levels	~1000%

	Plasma IGF-1 Levels	20%
	Blood Glucose	75%
	Plasma Insulin	25%
	Insulin Sensitivity	Increased
5	Glucose Tolerance	Impaired
	Food Consumption	52%
	Food Consumption (normalized to body weight)	137%
	Lifespan	126-140% *

10 \* varies with genetic background

RNA isolation For identification of differentially expressed genes, total RNA was isolated from the livers of 60 day old male GHR/BP -/- mice and +/+ controls using the RNA STAT-60 Total RNA/mRNA Isolation Reagent according to the manufacturer's instructions (Tel-Test, Friendswood, TX). For further characterization of the identified genes, expression in mice of other ages, in female mice, in other tissues, and in other mouse models, was considered as described below.

cdNA synthesis for subtraction library Prior to cdNA synthesis, a portion (50 µg) of RNA was further purified to remove small RNAs using the RNeasy Mini protocol for RNA clean up as instructed by the manufacturer (Qiagen Inc., Santa Clarita, CA). The cdNA was synthesized using 1 µg of total RNA from GHR/BP -/- and wildtype control mice using the SMART PCR cdNA Synthesis Kit according to the manufacturer's instructions (CLONTECH, Palo Alto, CA).

Generation of cdNA subtraction libraries Forward- and reverse-subtracted cdNA libraries were generated using the PCR-Select cdNA Subtraction Kit (CLONTECH, Palo Alto, CA) and the L-9, H-34 and H43 samples. The forward library (4) was of clones up-regulated in -/- mice compared to control

mice, The reverse library (5) was of clones down-regulated in -/- mice compared to control mice.

**Isolation of individual clones** After generating the cDNA subtraction libraries, the PCR product ends were made blunt by treatment with *Pfu* DNA polymerase (Stratagene, La Jolla, CA) and subcloned into a bacterial plasmid vector using the Zero Blunt TOPO PCR Cloning Kit as instructed by the manufacturer (Invitrogen Corp., Carlsbad, CA). Individual clones were obtained by plating on selective media.

**Screening by differential hybridization** cDNA arrays of clones from the forward and reverse subtracted libraries were screened with probes made from each library using the PCR-Select Differential Screening Kit according to the manufacturer's instructions (CLONTECH, Palo Alto, CA). Individual clones from each library were spotted in duplicate on each of two separate nylon membranes and hybridized with random-primed <sup>32</sup>P-labeled probes generated from each library pool using the Clontech PCR-Select Differential Screening kit. Potential differentially expressed clones were selected based on difference in signal between the two blots upon exposure to autoradiography film.

**Nucleotide sequence determination** Plasmid DNA from bacterial colonies carrying the differentially expressed cDNA inserts was isolated using the QIAprep Spin Miniprep Kit according to the manufacturer's instructions (Qiagen Inc., Santa Clarita, CA). Nucleotide sequences were determined by use of the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit with electrophoresis on the ABI PRISM 377 DNA Sequencer (PE Applied Biosystems, Foster City, CA.). Nucleotide sequences and predicted amino acid sequences were compared to public domain databases using the Blast 2.0 program (National Center for Biotechnology Information, National Institutes of Health).

**Northern analysis** Positive clones, identified by the differential hybridization screen, were used as probes in Northern hybridization analyses to confirm their differential expression. Total RNA (5-15 ug) isolated from GHR/BP -/- or wild-type control mice was resolved by agarose gel electrophoresis through a 1% agarose, 1 % formaldehyde denaturing gel, transferred to positively charged nylon membrane, and hybridized to a probe labeled with [32P] dCTP that was generated from the cDNA insert using the Random Primed DNA Labeling Kit (Roche, Palo Alto, CA), or to an asymmetric PCR amplified, digoxigenin (DIG) labeled probe synthesized from each clone of interest. In the latter case, blots were hybridized in DIG EasyHybe (Roche) and detection was performed following the manufacturer's guidelines (The DIG System User's Guide for Filter Hybridization, Roche).

**Database Searches** Nucleotide sequences and predicted amino acid sequences were compared to public domain databases using the Blast 2.0 program (National Center for Biotechnology Information, National Institutes of Health). Nucleotide sequences were displayed using ABI prism Edit View 1.0.1 (PE Applied Biosystems, Foster City, CA) or Vector NT 6.0 (Informax).

Nucleotide database searches were conducted with the then current version of BLASTN 2.0.12, see Altschul, et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res., 25:3389-3402 (1997). Searches employed the default parameters, unless otherwise stated.

For blastN searches, the default was the blastN matrix (1,-3), with gap penalties of 5 for existence and 2 for extension.

Protein database searches were conducted with the then-current version of BLAST X, see Altschul et al. (1997), supra. Searches employed the default parameters, unless otherwise stated. The scoring matrix was BLOSUM62, with gap costs of 11 for existence and 1 for extension. The standard low complexity filter was used.

"ref" indicates that NCBI's RefSeq is the source database. The identifier that follows is a RefSeq accession number, not a GenBank accession number. "RefSeq sequences are derived from GenBank and provide non-redundant curated data representing our current knowledge of known genes. Some records include additional sequence information that was never submitted to an archival database but is available in the literature. A small number of sequences are provided through collaboration; the underlying primary sequence data is available in GenBank, but may not be available in any one GenBank record. RefSeq sequences are not submitted primary sequences. RefSeq records are owned by NCBI and therefore can be updated as needed to maintain current annotation or to incorporate additional sequence information." See also <http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html>

It will be appreciated by those in the art that the exact results of a database search will change from day to day, as new sequences are added. Also, if you query with a longer version of the original sequence, the results will change. The results given here were obtained at one time and no guarantee is made that the exact same hits would be obtained in a search on the filing date. However, if an alignment between a particular query sequence and a particular database sequence is discussed, that alignment should not change (if the parameters and sequences remain unchanged).

**Construction and Synthesis of cDNA Libraries for Isolation of Longer cDNAs Using Differentially Expressed Partial cDNAs as Probes.** RNA was isolated from the liver of a 60 day old male GHR/BP homozygous knockout (-/-) animal using the RNA Stat-60 reagent and protocol (Tel-Test, Friendswood, TX). mRNA was isolated from the RNA sample using the Poly(A)Pure™ kit and protocol from Ambion (Austin, TX). cDNA synthesis was performed as per the manufacturer's protocol using ZAP Express cDNA Synthesis Kit (Stratagene, La Jolla, CA). The cDNA library was screened using previously isolated partial fragments obtained from the PCR-Select cDNA Subtraction Kit

(Clontech, Palo Alto, CA) and labeled with DIG using asymmetric PCR as described for Northern blot analysis.

#### Rapid Amplification of cDNA Ends (RACE). 5' Rapid

Amplification of cDNA Ends (RACE) was performed with the following primers in pursuit of the full-length clones for 5-9 and 5-61. RNA for the RACE reactions was isolated from a 60 day old GHR/BP homozygous knockout male mouse using the RNA Stat-60 reagent and protocol (Tel-Test, Friendswood, TX). RACE was performed as per protocol using the 5'/3' RACE Kit, 2<sup>nd</sup> Generation (Roche Applied Science, Penzberg, Germany). Annealing temperatures used for the 5-9 primers were 55 degrees Celsius and for the 5-61 primers were 65 degrees Celsius.

Primer	Sequence	SEQ ID NO:
5-9 SP1	GCTCTTTTCCTCTCACGGTAA	22
5-9 SP2	GGCTGCAAATGGTTCTGTAA	23
5-9 SP3	ACAGCCAGTAATGGACTCTTC	24
5-61 SP1	AGCTGTTTCAGGGCATTTC	25
5-61 SP2	TCAGCAAATGTCCACCAGTGCACA	26
5-61 SP3	GGAGTGAAGGCCATGACAGAGT	27

#### Derivation of Clone 5-43, 5-61 and 5-9 Sequences.

Clone 5-43. The sequence for Clone 5-43 was derived entirely from the clone obtained from the cDNA library. The full-length clone 5-43 was obtained by screening the cDNA library using a partial fragment corresponding to nucleotides 481 through 1068.

Clone 5-61. The sequence for Clone 5-61 was derived from a cDNA clone followed by 5' RACE. The cDNA clone was obtained by screening a cDNA library using a partial fragment corresponding to nucleotides 838 through 1377 that was previously isolated from a cDNA subtraction library.

Clone 5-9. The sequence for Clone 5-9 was derived from a partial cDNA clone followed by 5' RACE. The partial cDNA clone was obtained by differential hybridization of two cDNA subtraction libraries as described in the Methods section.

## Results

Of 192 total clones screened, ten clones appear to be differentially expressed, of which three are novel. The latter show increased expression in the livers of GHR/BP -/- mice. Each of these clones appears to be regulated in a unique fashion, thus they represent diverse gene regulation events that occur as a consequence of disrupted GH signaling.

### Additional Analysis of Clones 5-43, 5-9 and 5-61

#### *Expression in Alternative Mouse Models*

Using northern analysis, we compared the expression profiles of these clones in livers isolated from two additional mouse models of altered GH action. One line expressed the bovine GH (bGH) transgene, resulting in a giant phenotype. The other expressed a bGH antagonist (GHA) transgene, resulting in a dwarf phenotype. Northern analysis was also performed on multiple tissues of GHR/BP -/- mice and livers of GHR/BP -/- mice at various ages.

Total RNA was isolated from the livers of 60 day old mice; two male GHR/BP -/-, two female GHR/BP -/-, two male GHA, two male bGH mice and their respective controls. Hybridization was with each respective clone as probe. Clone 5-43 mRNA expression is evident in all liver samples tested and is elevated in the GHR/BP -/- male mice. A smaller transcript for clone 5-9 appears in all livers tested while a second, larger transcript appears only in the GHR/BP -/- mice. mRNA expression of clone 5-61 was only detected in GHR/BP -/- mice (both male and female).

#### *Expression in Other Tissues*

Total RNA was isolated from 60 day old male liver, kidney, muscle, heart, lung, spleen, brain, white adipose tissue (WAT), testis, intestine, stomach, and pancreas tissues. Clone 5-43 mRNA expression was detected at highest levels in liver, lung, WAT, intestine and stomach of both GHR/BP +/+ and -/- mice. Expression of the smaller



transcript of clone 5-9 mRNA is seen only in the livers of GHR/BP +/+ and -/- mice whereas the larger transcript is expressed in the liver, kidney and WAT of GHR/BP -/- mice. mRNA expression of clone 5-61 was limited to tissues of GHR/BP -/- mice with greatest expression detected in liver, kidney and WAT.

#### *Expression as a function of age*

Total RNA was isolated from livers of 2, 5, 12 and 24 month old female GHR/BP -/- and +/+ mice. Liver RNA was also isolated from these mice toward the end of their lifespan: 30 months for GHR/BP +/+ and 36 months for -/- mice. Hybridization was with each respective clone as probe. Clone 5-43 mRNA expression was detected in all samples of both GHR/BP +/+ and -/- mice. The smaller transcript for clone 5-9 mRNA was expressed in all samples of both GHR/BP +/+ and -/- mice while expression of the larger transcript was only apparent in the GHR/BP -/- mice. mRNA expression of clone 5-61 was detected throughout the lifespan of GHR/BP -/- mice and was not detected at any time point in +/+ mice. No significant age-dependent regulation of expression was evident for any of these clones.

#### *Conclusion*

Thus, Clone 5-43 mRNA is expressed in the liver of all three mouse models tested, in all tissues tested, as well as all non-transgenic controls. Highest expression was seen in the liver, kidney, intestine and brain. Clone 5-43 mRNA appears to be up-regulated in the livers of GHR/BP -/- males compared to controls. This up-regulation was not observed in GHR/BP -/- females.

Also, Clone 5-9 mRNA expression was seen in the liver of all three mouse models tested as well as all controls. A second larger transcript was detected in the GHR/BP -/- liver but was not present in other mouse models or controls. This larger transcript was also seen in kidney and white adipose tissue (WAT) of GHR/BP -/- mice but not controls. Interestingly, the smaller transcript was not detected in the kidney or WAT of the GHR/BP -/- mice.

The mRNA for our third clone, 5-61, was only detected in GHR/BP -/- males and females. Clone 5-61 mRNA appears to be expressed in the liver, kidney and WAT of GHR/BP -/- mice. In summary, these clones represent a diversity of gene regulation events that may be responsible for the physiological alterations that occur when the GHR/BP gene is disrupted.

In conclusion, we have identified three clones whose mRNA expression is differentially regulated in GHR/BP -/- mice. Clone 5-43 mRNA expression appears to be up-regulated in male GHR/BP -/- mice but is detected in every mouse model tested. The larger transcript for clone 5-9 is only present in the GHR/BP -/- mice. Similarly, clone 5-61 mRNA is only detected in GHR/BP -/- mice. Clone 5-43 mRNA, the larger transcript for 5-9, and 5-61 mRNA all appear to be expressed at relatively high levels in the liver, kidney and WAT of GHR/BP -/- mice, all significant organs of glucose metabolism and/or action.

## References Cited by Reference Number:

1. Semsei I. (2000) On the nature of aging. Mech Aging Dev 117:93-108.

2. Sohal, RS, Weindruch, R. (1998) Oxidative stress, caloric restriction, and aging. Science 273:59-63.

3. Finch, CE, Revkun, G. (2001) The genetics of aging. Annu. Rev. Genom. Hum. Genet. 2:435-462.

4. Roth, GS, Lasnikov, V, Lesnikov, M, Ingram, DK, Land, MA (2001) Dietary caloric restriction prevents the age-related decline in plasma melatonin levels of rhesus monkeys. J Clin Endocrinol Metab. 86:3292-5.

5. Roth GS, Lane MA, Ingram DK, Mattison JA, Elahi D, Tobin JD, Muller D, Metter EJ (2002) Biomarkers of caloric restriction may predict longevity in humans. Science. 297:811-813.

6. Walford RL, Mock D, Verdery R, MacCallum T. (2002) Calorie restriction in biosphere 2: alterations in physiologic, hematologic, hormonal, and biochemical parameters in humans restricted for a 2-year period. J Gerontol A Biol Sci Med Sci 57:211-24.

7. Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R. (1993) A C. elegans mutant that lives twice as long as wild type. Nature 366:461-464.

8. Lin, K, Dorman, JB, Rodan, A, Kenyon, C. (1997). daf-16: an HNF-3/Forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. Science 278, 1319-1322.

9. Clancy DJ, Gems D, Harshman LG, Oldham S, Stocker H, Hafen E, Leivers SJ, Partridge L. (2001) Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. Science 292:104-106.

10. Tatar, M, Bartke, A, Antebi. (2003) The endocrine regulation of aging by insulin-like signals. Science 299:1346-1351.

5

11. Tran H, Brunet A, Grenier JM, Datta SR, Fornace AJ Jr, DiStefano PS, Chiang LW, Greenberg ME. (2002) DNA repair pathway stimulated by the forkhead transcription factor FOXO3a through the Gadd45 protein. Science 296:530-534.

10

12. Ramaswamy S, Nakamura N, Sansal I, Bergeron L, Sellers WR. (2002) A novel mechanism of gene regulation and tumor suppression by the transcription factor FKHR. Cancer Cell 2002 2:81-91.

15

13. Hekimi, S, Guarente, L. (2003) Genetics and the specificity of the aging process. Science 299:1351-1354.

14. Brown-Borg, HM, Borg, KE, Meliska, CJ, Bartke, A. (1996) Dwarf mice and the aging process. Nature 384:33.

20

15. Flurkey K, Papaconstantinou J, Miller RA, Harrison DE. (2001) Lifespan extension and delayed immune and collagen aging in mutant mice with defects in growth hormone production. Proc Natl Acad Sci USA 98:6736-6741.

25

16. Zhou, Y, Xu, BC, Maheshwari, HG, He, L, Reed, M, Lozykowski, M, Okada, S, Cataldo, L, Coschigano, K, Wagner, TE, Baumann, G, Kopchick, JJ. (1997) A mammalian model for Laron syndrome produced by targeted disruption of the mouse growth hormone receptor/binding protein gene (the Laron mouse). Proc. Nat. Acad. Sci. USA 94:13215-13220.

30

17. Coschigano, K, Clemmons, D, Bellush, LL, Kopchick, JJ. (2000) Assessment of growth parameters and life-span of GHR/BP gene-disrupted mice. Endocrinology 141:2608-2613.

35

17a. Coschigano, KT, Holland, AN, Riders, ME, List, EO, Flyvberg, A, Kopchick, JJ, Deletion, but not antagonism, of

the mouse growth hormone receptor results in severely decreased body weights, insulin and IGF-1 levels and increased lifespan, *Endocrinology* (electronically published May 30, 2003 as doi:10.1210/en.2003-0374).

5

18. Holzenberger M, Dupont J, Ducos B, Leneuve P, Geloën A, Even PC, Cervera P, Le Bouc Y. (2003) IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature* 421:182-187.

10

19. Migliaccio E, Giorgio M, Mele S, Pelicci G, Reboldi P, Pandolfi PP, Lanfrancone L, Pelicci PG. (1999) The p66shc adaptor protein controls oxidative stress response and life span in mammals. *Nature* 402:309-313.

15

20. Bartke A, Wright JC, Mattison JA, Ingram DK, Miller RA, Roth GS. (2001) Extending the lifespan of long-lived mice. *Nature* 414:412.

21. Weindruch R, Kayo T, Lee CK, Prolla TA. (2002) Gene expression profiling of aging using DNA microarrays. *Mech Aging Dev* 123:177-193.

20

22. Lee CK, Allison DB, Brand J, Weindruch R, Prolla TA. (2002) Transcriptional profiles associated with aging and middle age-onset caloric restriction in mouse hearts. *Proc Natl Acad Sci USA* 99:14988-14993.

25

23. Prolla TA. (2002) DNA microarray analysis of the aging brain. *Chem Senses* 27:299-306.

#### *Additional References*

Coschigano KT, Riders ME, Holland AN, Kopchick JJ 2002 Altered growth hormone signaling in two lines of dwarf mice results in diet-induced obesity and hyperinsulinemia but not diabetes. 84<sup>th</sup> Annual Endocrine Society Meeting, San Francisco, CA, p. 561 (abstract P3-302)

30

Doi T, Striker LJ, Quaife C, Conti FG, Palmiter R, Behringer R, Brinster R, Striker GE 1988 Progressive glomerulosclerosis develops in transgenic mice chronically expressing growth hormone and growth hormone releasing factor but not in those expressing insulin like growth factor-1. *Am J Pathol* 131:398-403

35

List E, Coschigano KT, Kopchick JJ 2001 Growth Hormone Receptor/Binding Protein (GHR/BP) Knockout Mice: A 3-Year Update. Mol Genet Metab 73(1):1-10

5 Citation of documents herein is not intended as an admission  
that any of the documents cited herein is pertinent prior  
art, or an admission that the cited documents is considered  
material to the patentability of any of the claims of the  
present application. All statements as to the date or  
10 representation as to the contents of these documents is  
based on the information available to the applicant and does  
not constitute any admission as to the correctness of the  
dates or contents of these documents.

15 The appended claims are to be treated as a non-limiting  
recitation of preferred embodiments.

In addition to those set forth elsewhere, the following  
references are hereby incorporated by reference, in their  
most recent editions as of the time of filing of this  
application: Kay, Phage Display of Peptides and Proteins: A  
20 Laboratory Manual; the John Wiley and Sons Current Protocols  
series, including Ausubel, Current Protocols in Molecular  
Biology; Coligan, Current Protocols in Protein Science;  
Coligan, Current Protocols in Immunology; Current Protocols  
in Human Genetics; Current Protocols in Cytometry; Current  
25 Protocols in Pharmacology; Current Protocols in  
Neuroscience; Current Protocols in Cell Biology; Current  
Protocols in Toxicology; Current Protocols in Field  
Analytical Chemistry; Current Protocols in Nucleic Acid  
Chemistry; and Current Protocols in Human Genetics; and  
30 the following Cold Spring Harbor Laboratory publications:  
Sambrook, Molecular Cloning: A Laboratory Manual; Harlow,  
Antibodies: A Laboratory Manual; Manipulating the Mouse  
Embryo: A Laboratory Manual; Methods in Yeast Genetics: A  
Cold Spring Harbor Laboratory Course Manual; Drosophila  
35 Protocols; Imaging Neurons: A Laboratory Manual; Early  
Development of *Xenopus laevis*: A Laboratory Manual; Using  
Antibodies: A Laboratory Manual; At the Bench: A Laboratory  
Navigator; Cells: A Laboratory Manual; Methods in Yeast  
Genetics: A Laboratory Course Manual; Discovering Neurons:

The Experimental Basis of Neuroscience; Genome Analysis: A Laboratory Manual Series ; Laboratory DNA Science; Strategies for Protein Purification and Characterization: A Laboratory Course Manual; Genetic Analysis of Pathogenic Bacteria: A Laboratory Manual; PCR Primer: A Laboratory Manual; Methods in Plant Molecular Biology: A Laboratory Course Manual ; Manipulating the Mouse Embryo: A Laboratory Manual; Molecular Probes of the Nervous System; Experiments with Fission Yeast: A Laboratory Course Manual; A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria; DNA Science: A First Course in Recombinant DNA Technology; Methods in Yeast Genetics: A Laboratory Course Manual; Molecular Biology of Plants: A Laboratory Course Manual.

All references cited herein, including journal articles or abstracts, published, corresponding, prior or otherwise related U.S. or foreign patent applications, issued U.S. or foreign patents, or any other references, are entirely incorporated by reference herein, including all data, tables, figures, and text presented in the cited references. Additionally, the entire contents of the references cited within the references cited herein are also entirely incorporated by reference.

Reference to known method steps, conventional methods steps, known methods or conventional methods is not in any way an admission that any aspect, description or embodiment of the present invention is disclosed, taught or suggested in the relevant art.

The foregoing description of the specific embodiments will so fully reveal the general nature of the invention that others can, by applying knowledge within the skill of the art (including the contents of the references cited herein), readily modify and/or adapt for various applications such specific embodiments, without undue experimentation, without departing from the general concept of the present invention. Therefore, such adaptations and modifications are intended to be within the meaning and range of equivalents of the disclosed embodiments, based on the teaching and guidance presented herein. It is to be

understood that the phraseology or terminology herein is for the purpose of description and not of limitation, such that the terminology or phraseology of the present specification is to be interpreted by the skilled artisan in light of the teachings and guidance presented herein, in combination with the knowledge of one of ordinary skill in the art.

Any description of a class or range as being useful or preferred in the practice of the invention shall be deemed a description of any subclass (e.g., a disclosed class with one or more disclosed members omitted) or subrange contained therein, as well as a separate description of each individual member or value in said class or range.

The description of preferred embodiments individually shall be deemed a description of any possible combination of such preferred embodiments, except for combinations which are impossible (e.g., mutually exclusive choices for an element of the invention) or which are expressly excluded by this specification.

If an embodiment of this invention is disclosed in the prior art, the description of the invention shall be deemed to include the invention as herein disclosed with such embodiment excised.



## Table 1

## Clone 4-11

5 tatggcaaagatttgggtggagaaggccaagacctcagagattcagagccagggtcaaggcacactttgagaag  
acacacgagcagctgacaccccttgtcaggtcagcaggaactagtctggtgaacttcttcagcagtttaatg  
aacctcgaggagaaaccggctcctgctgagatgtgccaggcccagctcttcccaccccagctgctc  
cactggccaccgctagagcccctctccctaccttctgcctgttttct (SEQ ID NO:1)

## Table 2

## Clone 4-29

10 acgcgggggaattccaattctccaggcatacgcgtataacaactcggataaccattgttagttaatcagacta  
taggcaataatcacactataaataatccacctataacttctctgttaaccaacaccggaatgcctaaagga  
aagatccaaaaagataaaaaggaactcgnnaaacaagaaccccgctgtttacaaaaacatcacctctagca  
ttacaagtatttagaggcactgcctgcccagtgactaaagtttaacggccgcggtatcctgaccgtgcaaagg  
tagcataatcacttgttccttaattagggacnagcagaacggctaaacgagggtccaactgtctctttatct  
15 ttaatcagtgaattgacctttcagtgaaagggtgaaatataataaagacgagaagaccctatggagct  
taaattatataacttatctattttaatttattaaacctaatggcccaaaaactatagtataagtttgaaattt  
cggttgggggtgacctcggagaataaaaaatcctccgaatgattataaccttagacttacaagtcaaagtaaaa  
tcaacatatcttattgaccagatatattttgatcaacggaccaagttaccctagggataacagcgcaatcc  
tatttaagagttcatatcgacaattaggggtttacgacctcgatgttggtcaggacatcccaatgggtgtaga  
20 agctattaatggttcgtttgttcaacgattaaagtoctacgtgatctgagttcagaccggagcaatccaggt  
cggtttctatctatttacgatttctcccagtagctcggccgcgaccacgctaagggcgaattccagcacact  
g (SEQ ID NO:2)

## Table 3

## Clone 4-97

25 acgcggggggtggaaaggcgaccatgcttctctgctctgacagttgatgaaacatatgtcccaagagttta  
aagctgagaccttcaccttcactctgatctctgcacacttccagagaaggagaaacagattaagaaacaaa  
cggctcttctgctgagctggtgaagcacaagcccaaggctacagcggagcaactgaagactgtcatggatgact  
ttgcacagttcctggatacatgttgcaaggctgctgacaaggacacctgcttctcgactgaggggtccaaacc  
30 ttgtcactagatgcaaagacgccttagcctaaacacatcacaaaccacaaccttctcagggtaccctgagaaa  
aaaagacatgaagactcaggactcatc (SEQ ID NO:3)

## Table 4

## Clone 4-130

35 acgcgggggcaatcggttcacctcctcttccctccacgaaacaggatcaaacaaccaacaggattaaactcaga  
tgcagataaaaattccatttcacccccactatacaatcaaagatatcctaggtatcctaatacatattctta  
tctcataaccctagttattttttccagacatactaggagacccagacaactacataaccagctaattccact  
aaacacccccacccatattaaaccgaatgatatttcttatttgcatagccattctacgctcaatccccaa  
taaactaggaggtgtcctagccttaattcttatctatcctaatttttagccctaatacctttccttcatactc  
40 aaagcaacgaagcctaataattccgccaatcacacaaattttgtagctcgcccgccgaccacgctaagggcga  
attctgcagatatccatcacactggcgccgctcgagcatgcatctagagggcccaattcgccctatagtga  
gtcgtattacaattcactnggcccgtcgtttacaacgctentgactgggaaaacccctggcggttaccaaattaa  
tcngcctttgnaaacatccccctttcgccagctggncgnaatagcgaanaaggccccacccnacgcctnc  
caacagttnncccanctataacgtaacgnatttnanggtttacncctanaaaaaaaagagccg (SEQ ID  
45 NO:4)

## Table 5

## Clone 5-105

taaaaaattttttgggcacagctcggtttctagtaacgncgccagtggtgctggaattcgcccttagcgtgg  
 tcgcgccgaggtacgcgggggagtgtaggcactataacctgaaagacgtggctcctgacaggaggacaattc  
 tattccctaccaaaatgaagctgctgctgccgctgcttctgctcctgtgtttggaactgacttttagtctgtn  
 5 tncatgcagaagaatctagttctatggaaaggaactttaatgtagaacagattagtaggggtatttggtttct  
 attgctgaagcctctgatgaaagagaaaagatagaagaacatggcagcatgagagcttttgtggaaaacatc  
 actgtcttggagaattccttagtctttaaattccatttaattgtaaataagagtgccacgaaatgactgcg  
 attggtgaacaaacagaaaaggctggcatatattatatgaactatgatggattcaatacatttagtatactt  
 aagacagactatgataattatattatgattctctcattaacaaaaaggatgggaaaaccttcccagcntgat  
 10 ggagctctatggcncganaaccanatttngagtttaagacatcaaaggaaaagttttgcaaaactatgccga  
 ggagcatggaatcattagaaaaaaatatcattngacctaaccaangtccatccgctggcctcntaggccccc  
 aaaatgnaanaatggggccganccctccnggtggggcaaaattcanttanaacnangaggantgtnantngaaa  
 ntttttncaagactccactntccttcccttcaantccntgccaagnttnngatcc (SEQ ID NO:5)

## Table 6

## Clone 5-38

acgcggggagtgtagccacgatcacaagaaagacatggctcctgacagacagacaatcctattccctacaaaa  
 atgaagatgctgctgctgctgtgtttgggactgaccctagtctgtgtccatgcagaagaagctagttctacg  
 ggaaggaactttaatgtagaaaagattaatggggaatggcatactattatcctggcctctgacaaaagagaa  
 20 aagatagaagttaatggcaactttagactttttctggagcaaattccatgtcttggagaattccttagttctt  
 aaagtccatactgtaagagatgaagagtgtccgaattatctatgggtgctgacaaaacagaaaaggctggt  
 gaatattctgtgacgtatgatggattcaatacatttactatacctaagacagactatgataactttcttatg  
 gctcatctcattaacgaaaaggatggggaaaccttccagctgatggggctctatggccgagaaccagatttg  
 aggtcagacatcaaggaaagggtttgcacaactatgtgaggagcatggaatccttagagaaaatatcattgac  
 25 ctatccaatgccaatcgctgcctccaggcccgagaatgaagaatggcctgagcctccagtgttgagtggaca  
 cttctcaccaggactccaccatcatcccttccctatccatacagcatccccagtataaattctgtgatctgca  
 ttctatcctgtctcactgagaagtccaattccagtctatccacatgttacctaggatacctcatcaagaatc  
 aaagacttctttaaatttctctttgatatacccatgacaattttttnatgaatttcttcttcttctgttca  
 ataaatgattacccaaaaaaaaaaaaaaaaaaaaaaaaaaaaa (SEQ ID NO:6)

## Table 7

## Clone 5-41

gagtatccaagaaggagagtgatcgagtcctcgctatggagcatccccaactgcagccttactggcagga  
 cacggaaaggagagcggtgctctggctaacacagaacctttacgtttctgtgtctccactgtctctctgaa  
 35 cctccatgggtgctgcaagttaccagtggctaataagggtcccgactcccatcccatgctctatttaaaatgat  
 gatgtttggttctagactccat (SEQ ID NO:7)

## Table 8

## Clone 5-43

ctggagctcgcgccctgcaggaagcgaggagagcgcggtatccgagcgcccgcgggcggtcttctctgcgcgc  
 tcaggaccgcaccatgctcgcgctccgaagcggcctaaggacagcgctggcccccggggttctgacgcctca  
 ggtatgttcaccttttctacaggccccaggcaaaagcaatggcacattctatgaattccgcacctattttct  
 caagccctcaaagacaaatgagttcctggaaaattttaagaacagtgttcaccttcggacagctcactctga  
 aatgattggctattggactgtagaatttggaggcagaacgaacagagttttccatatttggaaagtatgataa  
 45 ttttgctcatcgaactgcagtccgcaaagccttggccaaggataaggaatggcaggaacgattcctcattcc  
 aaatttggctttcattgataaacaagaggttgagattacctacttggtaccatgggtgcaaaataggaacgcc  
 tcccaaggaaggagtctatgaactggctacctttcagatgaagcctggcgcccgagctctgtggggtaatgc

attttaaaggcagtaaagtcccatgtggagcttggctactctacactagttggtgtttccacaccgaata  
 tggagccctcaacagagttcatgttctgtggtggaacgagagtgacagacagccgagcagctgggagacactg  
 gtctcacgaggaccctagagtgggtggtgctgttctgggagagtggtcagttacctagagctccagcagaatac  
 gttcctgattccaacatcattttcaccactgaagtagtttctaccagcccagagcatcccatcacagagg  
 5 tgagctgtatccacagcgggcacttctgtctccaagatggagtcgcccagttcattgtctgaaggacgggg  
 ctgagctattacactgtgacctctgcatcttgaaggctacctgagtcatttactacgagtagttcaggaaat  
 gtttctgttcattttcataactactggatatcagaaagcattgttactccgtgccttggtacttctattgttt  
 cacaatatcttgcttctgagatttttatgactaagtcttgtattcaaacatgattttgaattttttccct  
 caaagtatctgtaaattaaataaatacaatgtctgcctttttatcatgttcatttttctgaaatttaaaac  
 10 tattttttaccagtaatcctttctttgtgataataagtgggatttatagaaatattaaatctagtagtttgtc  
 ttatacttttagaaaagtatttaattatgggtttcacattcattattaagtcaaaattggaaaataatatatcc  
 ttgagacttggtgtataatacgggtcttagtggtgtttgttttaattgcttgactacttttaactttacctt  
 gaattactcataaatattatgagacacttattttatttcaatgagaaaatggaaatctctgacctataaaaag  
 ctttcacattcaaaaaaaaaaaaaaaaaaaaaa (SEQ ID NO:8)

15

Translated coding sequence, bases 86-826 of SEQ ID NO:8:

MLALRSGRLTALAPRVLTQVCSFPATGPRQSNNGTFYEFRTYFLKPSKTNEFLENFKNSVHLRTAHSEMIGY  
 WTVFEGGRTNRVFIHWKYDNFAHRTAVRKALAKDKEWQERFLIPNLAFIDKQVEVEITYLVPWCKIGTPPKKG  
 20 VYELATFQMKPGGPALWGNFAFKRAVNAHVELGYSTLVGVFHTTEYGALNRVHVLWNNESADSRAAGRHWSHED  
 PRVVAAVRESVSYLESQNTFLIPTSFSPK\* (SEQ ID NO:12)

Table 9

Clone 5-61

25 gcgcaccgct cggtcgggcc cggaaagctt ctgcgagaag ctgtcgcggg tgggctcgga  
 gattgctagt ccctgggttg taccagtctt gagtctcggt gagactcggc gacgaaggtc  
 attccccttt aaaaaaaaaat atttgtttta attcatccga gttgttccgt ggccttttcc  
 tctccagaaa catgattctt tgcttaaat gacttatata ttaaaggagt gaaaatgcag  
 acagaactgt tcaaacgcca aggtccttg gcagtatcct gtttttatac gtatatccta  
 30 tctttgttca aattaatcgt gattgggtgg gctggatggg tgcttcctca gtctttatta  
 ggatttcac attgcttgaa agaaaccagt ctctacctg ataaaaatgt gaatagattg  
 ttagtcataa acttaagtgt agtaagaata gagactgtag ctttaaaaga gatgtgacgg  
 gaggtacacc tgtagcacag gtagggcagg actggtttcc aggcaactgt caggggtgaa  
 gaaagggtgag gtataaattg cttttttttt ttttttcac ctggaaaaaa aaagaaaaaa  
 35 gaaaagttct acagttttac acttccttaa gtctatacag gtatagttaa tatgcatcta  
 ttaagactgg gacatttttg agttcactgt ggtggctctt gcctttaaat tttttttttt  
 cttgaaacat ggagatttta agagagattt taccaaaata gagaccaaag ataacttaaa  
 tacctttatc agtaaaagag caaattttgt aaggaataag ctggtttcgt actaacctgt  
 caactctggc ggatattatc aggtttgaga ggaaatacat gaagatgtt aaaaaggctg  
 40 ggcactctgt catggccttc actccctctc ctgtgcaactg gtggacattt gctgatcagt  
 ggtaaaaatc tcaaactgta aatctgtatc ttggcaaata aatcagatgg taaactcctt  
 aagggctgga aaatgccctg aacagctctt attttacaca gccttgaata tagttcatgt  
 tgattaaata catgcaggaa agcatgggtt acatttttaag gcttaacagg attgggtcaa  
 gaactgataa ggtttggaag gaaaatctgt tgttccttgt taatagctgc tatgaaatgt  
 45 ctggatttct catatttttg agaggtagaa ttattgtttt ttaatgagtt ggatttccta  
 ctttgggttc atagtagata ctggtgagcc ttttgatga ctttaattatg tcgtagtctc  
 taggtgtttt atgttttatg aaactgagtc tttattgcca cttctccaaa gctgtactag

atgaatactg cctttcagct catactaata gtagctttta aaagcaagtt ccctaaatta  
 cagggtgcacc acaccagctt gcaactgacc tgattgtcag catacctaca tcagggttgc  
 ttggaaaaggc aagcagagga gggaaaagggtg gtgcccagggt taattctagg aatttgaatg  
 tccttggtc cctctaccag agaagcagct ctgctcctaa tctccataaa tatgaactgg  
 5 tgectcacca tccagatgca cgtagatagc atcctttgca gacatttggc caggacatgg  
 gaaagctaata gaacagtaac tagataaaaa ccatccacct gtggtattct ggcattctc  
 ccacagttta aacttgatgt ggataaaact ggtgaacttc agggctctaaa ctataaatta  
 ctagcttaac aggatgggtg ttaagaaaaa aaagtgtggg attcaaactt caaaagacaa  
 gattaggagt cttcattgat attttagtca attcctttgt actgtaaaat tggagtggga  
 10 ctattgcaaa gtgactgatt tcccaggcct tgctgaatat atctttaaaa tgtggaattc  
 gagccctttt cagtttgga ataggatggg tagcttcctg ctcaaaagct cttacaaccc  
 atagctaaga agaaacttca caagttacag attccttaatt cacatgttct tttctggcct  
 cgtctcatgg ctggctatgt tatttaagaa atcagttctt agcactgtgg cacacacctt  
 taatcctagc agtagcacta ggcagaggca ggcagatctc taagttctaa ggccaacctg  
 15 gtctacagag tgatgtctgg acagctggag ctacaaagag aaagcttgtc tggaaaacca  
 aaacaaaaaa ctataaaaa aaaggaaatc agttgtgaat gtcgcctatc tgtggaggcc  
 tcccctgacc acctgtgtaa ataagctctc tgccctcctta tccctgccatt gtgccctgct  
 catttccttg agagataacc ccatgtgaaa ttatcctcat cctcgtaga tggcctcatc  
 agccaacaac catttttctt catctgatga atctagtcca gaagaagcat tcctatgtat  
 20 ttgtgaaaat aaatgtttgc tttcttagag tttgacctga agaaagctga gttccagttc  
 ctgtcattgc aggtagctag cacataggtc ttatgtagcc ctccttttta tagaaacatg  
 attttaggtc agcaagaata ttaggagaac caaggctatt gcctaaaatg acgatagaaa  
 ggaggagtgg gtgaagagag cgctggctgg ctgacagcgc agaacagccc aggtaggatg  
 cttcagtcag cagggtgtct caactcactt catttcctta aagcaaaaat ggagattata  
 25 attcctgcct cttacaggga tgtgtagata gtatataaga gagagctttg taattttagt  
 agtgctttac agataaaaaga atcctattac tgaatttcag gattttttta aaaattattt  
 atttttattt atatgaatac actgaagaca gctacagtgt acttacatat aataataaat  
 aaatctttta aaaaaaaaaa aaaaaaaaaa a (SEQ ID NO: 9)

### 30 Clone 5-61 Open Reading Frames

>lcl|Sequence 2 ORF:235..477 Frame +1

MQTELFKRQGS LAVSCFYTYILSLFKLIVIGGAGWVLPQSL LGFHHCLKETSLLPDKNVNRL LVINLSVVR I  
 ETVALKEM\* (SEQ ID NO: 13)

35 >lcl|Sequence 7 ORF:463..657 Frame -2

MHIHYTCIDLRKCKTVELFFFLFFSRMKKKKAIYTS PFFTPDSCL ETSPALPVLQVYLP SHLF\* (SEQ  
 ID NO: 14)

40 >lcl|Sequence 1 ORF:1077..1244 Frame +3

MLIKYMQESMVYILRLNRIGPRTDKVWKENLLFLVNSCYEMSGFLIFLRGRIIVF\* (SEQ ID NO: 15)

45 >lcl|Sequence 8 ORF:1491..1652 Frame -3

MLSTCIWMVRHQFIFMEIRSRAASLVEGARDIQIPRITLGTTFPSSACLSKAT\* (SEQ ID NO: 16)

>lcl|Sequence 6 ORF:776..913 Frame -1

MTECPAFLNIFMYFLSNLIISARVDRLVRNQLIPYKICSFTDKGI\* (SEQ ID NO: 17)

5 >lcl|Sequence 3 ORF:1558..1689 Frame +1

MSLAPSTREAALLLISINMNWCLTIQMHVDSILCRHLARTWES\* (SEQ ID NO: 18)

>lcl|Sequence 4 ORF:2084..2203 Frame +2

10 MFFSGLVSWLAMLFKKSVLSTVAHTFNPSSSTRQRQADL\* (SEQ ID NO: 19)

>lcl|Sequence 5 ORF:617..730 Frame -1

15 MFQEKKKFKGKSHHSELKNVPVLIDAYSLLYRLKEV\* (SEQ ID NO: 20)

# Table 10

## Clone 5-9

20 cc agatacattc agatagccaa aggtttctga ccgacctccg gaatgatatt gagaagaaaa  
taggcttcga tgctatcatg agggttcggg ccagcacagg tttcagggcc actgatttct  
ttggtggaat attcatgaac aacaccacag atgtggaaat ggcggccatc gactgcgata  
aagccgtgac agtggagttc aagcacgatg acaagctcag tgaagatgtc ggggccctga  
tccagtgtgc tgtgctctac acgacctca gtggtcagcg aagactccgg attcacaatc  
25 tcgccttaaa ctgcagcacc cagctagcgg atctgtacaa gagctgtgaa actgatgcc  
ttatcaactt ctttgccaag tcagctttta aagcagtttt gaaccagccc ttgaaggcca  
tccgggaaat tctagtcaat cagactgccc atatgctggc gtgttaccgg aagcactgtg  
cgagtcctgc tgcagcaagc cagcttatac tgccagattc catgaagggtg ctgcctgtgt  
acatgaacag tttgttgaaa aactgtgtgc tgctcagcag atcggagatc tcccctgatg  
30 agcgggctta ccagcggcag ctgggtcatga ccatgggcgt ggctgactcc cagctcttct  
tctatccact gtcctgccca atccacacgt tagatgtgaa gactgcagcg ctgccaccag  
ccgtccgctg ctccgagtc cgactctctg aagaaggaat attcctgctg gctaacggtc  
tgaacatggt cttatggttt ggagtgggca gccaccaga gctgattcag ggaatattta  
atgtgccatc gtttgacat atcaacacag atatgacatc gctgcctgag gtgggaagtc  
35 cacactctca acaactcaga atgataatga ataatatcca gcaaaagaag ccataactcaa  
tgaagctcat cgtgggttaa cagcgagagc agcgggagat ggctttccga cagtccctgg  
tggaagataa agggctctat ggaggatctt cgtacgtgga tttcctctgc tgtgttcaca  
aggagatctg tcaactgctt aactgagccc tcggctctcc ctctgaggtc acattcctga  
ggagacagtc tccaacttgg tgctaattt tctgatgata gtaggctagt ttaatttct  
40 tgatcacttt caggacagct tgcctggaca gcacacagac cttcagctcg gactgctcag  
gtggcaagcc cgtggaggtg tgggggtact gtaaaggga gagtccatta ctggctgtcc  
agcttcta at ttctaaagct gacacaattt gcatttcata aatacatagt ttacagaacc  
atttgcagcc tcaattttct ttgtgcttgt tatataaatt atttttcaaa ctgtatagat  
tcatgataaa attatcttgg ttctctctc ttttaccgtg agaggaaaag acttcatttt  
45 tatcttatac ctaatttttt taaaaccccg tgatgacatt aaacagaatt tttttccaa  
cataagatct gcacagcaga cttttaaaaa gcctgaaatc tgtctggtag aacaatttgt  
gttctccttt tttcataatt atatattgtc tgtttaaaat atttccagt tgttttgtct

ataaaacggt atctatattc ttaatggttc tttgtacaat tttgatggtt tctacctgta  
 tataatggat cttaaccaat atcaataaat tacttcatgt aaaaaaaaaa aaaaaa (SEQ ID  
 NO: 10)

5 5-9 larger transcript ORF:

MRVRTSTGFRATDFFGGIFMNNTTDVEMAAIDCDKAVTVEFKHDDKLSDEVGALIQCAVLYTTISGQRRRLRI  
 HNLALNCSTQLADLYKSCETDALINFFAKSAFKAVLNQPLKAIREILVNQTAHMLACYRKHCASPSAASQLI  
 LPDSMKVLPVYMNSLLKNCVLLSRSEISPDERAYQRQLVMTMGVADSQLFFYPILLPIHTLDVKSALPPAV  
 RCSESRLSEEGIFLLANGLNMFLWFGVGSPPELIQGIFNVPSFAHINTDMTSLPEVGSPHSQQLRMIMNNIQ  
 10 QKKPYSMKLIVVKQREQREMAFRQFLVEDKGLYGGSSYVDFLCCVHKEICQLLN\* (SEQ ID NO:21)

Table 11

Clone 5-138

aaaaggatggggnnnccttccagntgatggggntttatggccgagaaccagntttgagttcagacatcaagg  
 15 aaagggtttgcaaaaantntgtgaggagcatgggnccttananaaaatatcattgacctatccaatnccaatn  
 gctgcctccngggcccgagaatgaanaatggcctgagcctccaggtgggcaancnccnaanagagcaaggagg  
 ggggttggtcgacatggagaggcccttcccaaagtattaatgttggtgcccccaaattncattaaaaaatagtt  
 ngataaaanantntttagtaggaagcttgagaaaaaaaaaaaaaaaaaaaaaa (SEQ ID NO:11)

20 In the above Tables, "N" denotes "unknown".

## CLONE DESCRIPTIONS

4-11 is not full length but the mouse homolog is. The human homolog is not full length.

5

4-29 is not full length but it is matching with a portion of the mitochondrial genome, specifically the 16s ribosomal RNA. There is no corresponding protein match for this reason.

10

4-97 is not full length but the mouse and human homologs are.

15

4-130 is not full length but the mouse and human protein homologs are.

5-105 is not full length but the mouse homolog is and the potential human protein homolog is not.

20

5-38 is full length. The potential human homolog is not full length.

25

5-41 is not full length but the mouse and human homologs are. \*The human and mouse protein homologs are listed by inference since the 5-41 sequence is predominantly in the 3' non-coding region.

30

5-43 is a novel 1542 bp full length cDNA isolated from a GHR/BP -/- liver cDNA library. This cDNA encodes a 247 amino acid protein of unknown function.

5-61 is a novel partial clone.

35

5-9 is a novel clone. Clone 5-9 shows two different sized transcripts via Northern analysis. The larger transcript only shows up in the GHR/BP -/- mice. The databases show several mRNA sequences that increase in size, all at the 5' end. The 2766 bp mRNA BC030852 is used here as a reference

for nucleotide numbers. Our original clone fragment matches the nucleotide sequence of BC030852 from 2367 to 2683. We believe we have isolated the full length transcript for the larger and smaller transcript. From GHR/BP +/- liver RNA we isolated a 1321 bp cDNA clone encoding a 194 amino acid protein. This clone matches the nucleotide sequence from 1430 to 2766 on BC030852. From GHR/BP homozygous knockout (-/-) kidney RNA we have isolated a 1782 bp cDNA clone encoding a 342 amino acid protein. This clone matches the nucleotide sequence from 969 to 2766 on BC030852. The following sequences showing homology to clone 5-9 have been reported in GenBank. Their only difference is that they get progressively shorter from their 5' end: BC030852, BC038038, BC049257, BC032094, BC026624, AK009425.

5-138 is full length. The potential human homolog is not full length.



### *Introduction to Master Table*

Col. 1: The internal designation for the clone. The sequences for the clones appear in tables 1-11.

Col. 2: There are three pieces of information here: (1) The database accession number for the mouse gene "corresponding" to the clone as determined by database searching, (2) in parentheses, the E value for the alignment of the clone sequence to the mouse gene. It is the expected number of matches with the same or better alignment score that would have occurred through chance. The lower the E value, the more statistically significant the alignment. (3) the database accession number for the mouse protein corresponding to the mouse gene above.

Col. 3. "U/F". "U" means an unfavorable differential pattern of expression, "F", a favorable one. "F" means -/- expression greater than +/+ (control) expression. Master table 1 is divided into subtables 1A ("F") and 1B ("U") on the basis of this entry.

Col. 4: A human protein deemed to correspond to the clone, identified by database accession number and by name. Note that more than one human protein may be so identified. The human proteins are listed in order of correspondence to the clone, from most to least closely corresponding.

Col. 5: The E value for the alignment of the query sequence set forth in col. 6 to the human protein set forth in col. 4. There is one entry for each human protein in col. 4.

Col. 6. The database accession number of the corresponding human gene. There is one entry for each human protein in col. 4.

Master Table 1

Clone	Mouse Gene; (E Value); Protein	U	Human Protein: Human Protein Name	E Value	Human Gene
<b>Subtable 1A: Favorable Genes/Proteins</b>					
4-11	NM_013474. 1 (e-144) NP_038502. 1	F	1L6LA Chain A, Structures Of Apolipoprotein A-II And A Lipid Surrogate Complex Provide Insights Into Apolipoprotein-Lipid Interactions gb AAB32200.1  apolipoprotein D, apoD ref NP_001634.1 apolipoprotein A-II precursor	7e-013 7e-013 7e-013	None avail. None avail. X02619
4-29	AB042432 (0.0) none	F	AAK50430. Humanin	1e-076	AY029066
4-97	XM_132149. 2 (0.0) XP_132149.1	F	gb AAN17825.1  serum albumin gb AAH14308.1 AAH14308 Unknown (protein for IMAGE:3934797) gb AAF71067.1 AF116721_45 PRO1708 gb AAF22034.1 AF118094_29 PRO2044 gb AAH35969.1  similar to serum albumin precursor	4e-041 5e-041 5e-041 5e-041 5e-041 3e-041	AF542069 BC014308 AF116645 AF118090 BC035969 AF465942

	(0.0) CAA40926.1		gb AAN14559.1  cytochrome b			
5-41	NM_008063.1 (e-131) NP_032089	F	NP_001458 glucose-6-phosphatase, transport (glucose-6-phosphate) protein 1; Glucose-6-phosphate transporter-1	7e-004		NM_001467
5-43	NM_025623 (0.0) NP_079899.1	F	pir  T17302 hypothetical protein DKFZp564D177.1	e-123		Non avail.
			gb AAF28977.1 HSPC299	e-121		AF161417.1
			NP_056284 nipsnap homolog 3A	e-121		NM_015469.1
			BAB14777 unnamed protein product	e-121		AK024015.1
			NP_060846 nipsnap homolog 3B	e-115		NM_018376.2
			BAA92101 unnamed protein product	e-114		AK002137.1
5-61	AL607084.14 (0.0) No Sig	F	No Significant homology found			
5-9	larger transcript BAA34475		NP_055637 Sec24-related protein D; protein transport protein Sec24D; SEC24			
		F	(S. cerevisiae) related gene family, member D	0.0		NM_014822.1
	(0.0)		BAA34475 KIAA0755 protein	0.0		AB018298.1

AB018298.1	AAH35761 Sec24-related protein D	0.0	BC035761.1
	P53992 Protein transport protein Sec24C (SEC24-related protein C)		
		e-124	SEC24C
	NP_004913 SEC24-related protein C; protein transport protein SEC24C	e-124	NM_004922.2
	BAA07558 KIAA0079	e-110	D38555.1
	BAA91334 unnamed protein product	3e-56	AK000709.1

Subtable 1B: Unfavorable Genes/Proteins

5-105	BC019965.1 (0.0)	U	sp Q8WX39 MUPL_HUMAN Putative MUP-like lipocalin precursor	2e-015	None avail.
5-38	AAH19965.1 XM_135574.1 1 (0.0) 1MUP	U	sp Q8WX39 MUPL_HUMAN Putative MUP-like lipocalin precursor	1e-025	None avail.
5-138	BC037152.1 (e -120) pir  S10124	U	ref XP_071155.1 similar to major urinary protein 4 [Mus musculus] [Homo sapiens]	1e-004	XM_071155

## Notes

None avail. means that the protein sequence came from direct sequence of the protein, not via a DNA clone.

5-9 Clone nucleotide sequence is 3' of mouse CDS. Human protein identified by Blast P with mouse protein sequence.

4-29 Claiming human protein based on homology to 5' and 3' untranslated regions. (human e-value is from Blastn)

5-41 Clone nucleotide sequence is 3' of human CDS. (human protein and e-value is from Blastn)

5-105 Human peptide was sequenced, gene not cloned

5-38 Human peptide was sequenced, gene not cloned